

Molecular Mechanisms of Repulsive Guidance Molecule Signalling

Eleanor G. Healey

A thesis submitted in partial fulfilment of the requirements for the
degree of *Doctor of Philosophy*



Keble College
Nuffield Department of Medicine
Division of Structural Biology
University of Oxford

Trinity 2015

Molecular Mechanisms of Repulsive Guidance Molecule Signalling

Eleanor G. Healey

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

Keble College, Nuffield Department of Medicine and
Division of Structural Biology
University of Oxford, Trinity 2015

Abstract

Repulsive Guidance Molecules (RGMs) control fundamental and diverse cellular processes including axon guidance, immune cell regulation and systemic iron metabolism. RGM dysfunction has been linked to diseases such as multiple sclerosis, cancer and the iron overload disorder juvenile hemochromatosis (JHH). RGMs signal by binding directly to the transmembrane receptor Neogenin (NEO1) to trigger cytoskeleton rearrangements and subsequent axon repulsion. Additionally, RGMs are important activators of the essential developmental Bone Morphogenetic Protein (BMP) signalling pathway. RGM-activated BMP signalling is crucial for the regulation of iron metabolism and mutations in RGMs cause JHH. This thesis outlines structural and functional studies of the molecular mechanisms of RGM signalling. In Chapter 3, an analysis of an RGM-NEO1 crystal structure is presented. Combined with mutagenesis studies, analytical ultracentrifugation experiments and neurite outgrowth assays, this allowed a mechanism for RGM signalling through NEO1 to be proposed. RGM acts as a molecular staple, bringing together the juxtamembrane regions of two NEO1 receptors to activate downstream signalling. In Chapter 4, crystal structures of the N-terminal domains of all RGM family members in complex with the BMP2 ligand are presented. Together with biophysical and cellular assays these structures allowed an endocytosis-linked mechanism for RGM-activated BMP signalling to be proposed, which is dependent on the subcellular localisation of the BMP-receptors. The work outlined in both of these chapters also revealed a molecular rationale for the disease-causing mechanism of RGM-linked mutations. In Chapter 5, the crystal structure of the ternary BMP2-RGM-NEO1 complex is described along with super-resolution fluorescence microscopy data demonstrating BMP-induced clustering of RGM-NEO1 complexes in the membrane. In summary, this work sheds light on the molecular mechanisms of RGM signalling through NEO1 and BMPs, and demonstrates for the first time that RGM forms a structural bridge between these two fundamental signalling pathways.

Acknowledgements

I would like to begin by thanking my supervisor Christian Siebold for his unstoppable enthusiasm, brilliant ideas, encouragement and support throughout the four years of my DPhil. I also wish to thank him for giving me the opportunity to work in such a fantastic place as STRUBI, which is unique in its diverse range of equipment, expertise and people. Thanks to the rest of the Siebies; Ben, Ross, Sam, Eamon, Lea, Amalie, Christiane and Joy; and to Amber, Jonathan, Ester and Paul for making the lab such an enjoyable place to work, helping out with experiments, answering my questions and listening to my moaning.

I would particularly like to acknowledge Sergi Padilla-Parra for his help with the microscopy data collection and analysis, Yvonne Jones and Radu Aricescu for helpful discussions and ideas, Benjamin Bishop for his help with mammalian cell culture, microscopy and SPR, Ross Robinson for wet-lab guidance and MALS expertise, Jonathan Elegheert for his help with MALS and SAXS data collection and analysis, Robert Gilbert for his assistance with AUC, Christian Bell for the RGMB-NEO1 structure, cloning and SPR, Tom Walter and Karl Harlos for help with protein crystallisation and data collection, and Margaret Jones for invaluable lab support.

Thanks to friends throughout Oxford for making it a fun four years, particularly the Wizards of Oz and the 158 gang; special thanks to Liam, Annabel and Max for giving up lots of time for me. Finally, I would like to thank my parents, Mark and Lisa, for their love and support; without them, none of this would have been possible.

Technical Acknowledgements

This work was funded by the Wellcome Trust and Cancer Research UK.

I would like to acknowledge the following contributions to specific experimental data reported in this thesis:

- Sergi Padilla-Parra performed the analysis of the TIRF-dSTORM data, Figures 5.10, 5.11.
- Jonathan Elegheert carried out the minimal ensemble search analysis of the SAXS data, Figure 5.7, 5.8.
- Susan van Erp in the Pasterkamp laboratory (Utrecht) conducted the neurite outgrowth assays and data analysis, Figure 3.15.
- Christian Bell performed the eRGMB-NEO1_{FN56} crystallisation and structure solution, Figures 3.2-3.5; and the eRGMB-NEO1_{FN56} MALS analysis, Figure 3.6.
- Benjamin Bishop performed the SPR analysis of RGM-NEO1 binding in DDM, Figure 3.16c.

Abbreviations

ACVR2A	Activin receptor type II A
eACVR2A	ectodomain of Activin receptor type II A
AP2	adaptor protein 2
BISC	BMP-induced signalling complex
BMP	Bone Morphogenetic Protein
BMPR1A	Bone Morphogenetic Protein Receptor type IA
eBMPR1A	ectodomain of Bone Morphogenetic Protein Receptor type IA
BMPR2	Bone Morphogenetic Protein Receptor type II
eBMPR2	ectodomain of Bone Morphogenetic Protein Receptor type II
DCC	deleted in colorectal cancer
DRG	dorsal root ganglion
EAE	Experimental autoimmune encephalomyelitis
EEA1	early endosome antigen 1
FAK	focal adhesion kinase
FGF	fibroblast growth factor
FN	fibronectin
GAG	glycosaminoglycans
GDNF	glial cell-derived neurotrophic factor
GFR α 1	GDNF family receptor α 1
GPI	glycophosphatidylinositol
GTPase	guanosine-5'-triphosphate hydrolysing enzyme
HA	haemagglutinin
HH	Hereditary Hemochromatosis
Ig	Immunoglobulin
IL-6	Interleukin-6

JHH	Juvenile Hemochromatosis
KD	kinase domain
MS	multiple sclerosis
NCAM	neuronal cell adhesion molecule
NEO1	Neogenin
NEO1 _{FN56}	The FN5-FN6 domains of NEO1
PDGF	platelet derived growth factor
PD-1	programmed cell death protein 1
PD-L2	programmed death ligand 2
PFC	pre-formed receptor complex
RGM	Repulsive Guidance Molecule
eRGM	ectodomain of Repulsive Guidance Molecule
RGM _{ND}	N-terminal domain of Repulsive Guidance Molecule
RGM _{MB}	Middle domain of Repulsive Guidance Molecule
RGM _{CD}	C-terminal domain of Repulsive Guidance Molecule
RGM _{ΔN}	Repulsive Guidance Molecule lacking the N-terminal domain
ROCK	Rho-associated protein kinase
SMAD	similar to mothers against decapentaplegic
TER	transepithelial resistance
TGF β	Transforming Growth Factor β
TGF β R1	Transforming Growth Factor β receptor 1
TGF β R2	Transforming Growth Factor β receptor 2
vWfD	von Willebrand factor type D

Contents

1	Introduction	14
1.1	Development, morphogens and cell migration	15
1.1.1	Morphogen gradients	16
1.1.2	Overview of early development	18
1.1.3	The role of cell migration in brain development	24
1.2	The discovery of Repulsive Guidance Molecules	28
1.3	Functions of RGMs	34
1.3.1	RGMA	34
1.3.2	RGMB	38
1.3.3	RGMC	40
1.4	The role of RGMs in disease	42
1.4.1	Cancer	42

1.4.2	Multiple Sclerosis	44
1.4.3	Regenerative failure	45
1.4.4	Iron homeostasis disorders	45
1.5	RGM signalling through NEO1	47
1.6	RGM signalling through BMPs	53
1.7	The interplay between the NEO1 and BMP signalling pathways	58
1.8	Thesis outline	64
2	Materials and Methods	68
2.1	Molecular Biology	69
2.1.1	cDNA and expression vectors	69
2.1.2	Construct design and PCR	70
2.1.3	Restriction digests, ligations and transformations	71
2.1.4	Site-directed mutagenesis	71
2.2	Western Blotting	72
2.3	Protein expression and purification	73
2.3.1	Expression and purification of BMP2 from <i>E Coli</i>	73
2.3.2	Small-scale expression trials in mammalian cells	74
2.3.3	Expression and purification of secreted proteins from mam- malian cells	75

2.3.4	Co-immunoprecipitation of trans-membrane proteins . . .	76
2.4	Protein crystallisation and data collection	77
2.5	Basic principles of biomolecular crystallography	78
2.5.1	Protein crystal symmetry	78
2.5.2	X-ray diffraction	79
2.5.3	Bragg's Law	81
2.5.4	Reciprocal space	81
2.5.5	Structure factors	83
2.5.6	Reconstructing electron density	83
2.6	Structure solution pipeline	85
2.6.1	Data processing	85
2.6.2	Molecular replacement	89
2.6.3	Model building and refinement	91
2.7	Surface Plasmon Resonance (SPR)	96
2.8	Small-angle X-ray scattering (SAXS)	98
2.8.1	Basic principles of SAXS	98
2.8.2	SAXS data collection and processing	100
2.9	Multiangle light scattering	102
2.10	Luciferase reporter assay	102

2.11	Fluorescence microscopy	103
2.11.1	Principles of TIRF-dSTORM	104
2.11.2	Principles of FRET-FLIM	105
2.11.3	Sample preparation	107
2.11.4	dSTORM imaging	108
2.11.5	dSTORM data processing	108
2.11.6	FRET-FLIM imaging	109
2.11.7	FRET-FLIM Image Analysis	110
2.11.8	Live-cell imaging of fluorescently-labelled BMP2	110
2.12	Analytical ultracentrifugation	111
2.13	Neurite growth assays	112
3	The Repulsive Guidance Molecule (RGM)-Neogenin signalling hub.	114
3.1	Abstract	115
3.2	Introduction	115
3.2.1	RGM binds NEO1 at the FN5-6 domains	115
3.2.2	The structure of the RGM-NEO1 complex	117
3.2.3	The structure of the RGM middle domain	120
3.2.4	A higher order RGMB-NEO1 complex?	124

3.2.5	Chapter outline	125
3.3	Results	126
3.3.1	JHH-linked mutations abolish the secretion of RGMC . . .	126
3.3.2	A structural analysis of the 2:2 eRGMB-NEO1 _{FN56} complex	129
3.3.3	AUC analysis of eRGMB and NEO1 _{FN56}	131
3.3.4	Design, expression and purification of a site-2 mutant	133
3.3.5	The observed 2:2 eRGMB-NEO1 _{FN56} complex exists in solution	136
3.3.6	Formation of the 2:2 complex is pH-dependent	138
3.3.7	The 2:2 eRGMB-NEO1 _{FN56} complex is important for sig- nalling	140
3.3.8	NEO1 pre-clusters in the plasma membrane	143
3.3.9	RGM binding rigidifies the structure of the NEO1 receptor	145
3.4	Discussion	147
3.4.1	A potential model for RGM signalling through NEO1 . . .	147
3.4.2	A pH-dependent mechanism of signalling regulation?	149
3.4.3	The role of the site-2 interface	149
3.4.4	NEO1 as a central target for RGMs and Netrin-1	150

3.4.5	Insights into the JHH-causing mechanisms of RGM C mutations	151
3.5	Summary	152
4	Molecular mechanisms of RGM-activated BMP signalling	153
4.1	Abstract	154
4.2	Introduction	154
4.2.1	Structural studies of BMP signalling	155
4.2.2	RGMs as activators of BMP signalling	161
4.2.3	JHH-linked mutations may identify the BMP binding site on RGMs	162
4.2.4	Chapter outline	162
4.3	Results	164
4.3.1	Expression and purification of RGM C _{ND} and BMP2	164
4.3.2	The structure of the RGM C _{ND} -BMP2 complex	168
4.3.3	The BMP2-binding mode is conserved across the RGM family	175
4.3.4	Validation of the RGM _{ND} -BMP2 interface	177
4.3.5	The structure of the eBMPR1A-BMP2 complex reveals new features	182
4.3.6	RGM competes with BMPR1A for BMP2-binding	184

4.3.7	The RGM-BMP2 interaction is pH-dependent	188
4.3.8	RGMB and BMPR1A have different effects on cellular BMP signalling	190
4.3.9	Investigating internalisation of the BMP ligand	193
4.3.10	Tracking the subcellular localisation of BMP-receptor com- plexes	196
4.3.11	Time-resolved analysis of BMP receptor FRET	202
4.4	Discussion	204
4.4.1	An analysis of RGM interactions	204
4.4.2	A model for RGM-activated BMP signalling	206
4.4.3	A cellular approach to analysing BMP signalling	210
4.4.4	Implications of the RGM-BMP2 complex structure for JHH	212
4.5	Summary	213
5	RGM is a structural bridge between BMP and NEO1 signalling.	215
5.1	Abstract	216
5.2	Introduction	216
5.2.1	NEO1 and BMP signalling are physiologically linked . . .	217
5.2.2	RGM plausibly forms a physical link between NEO1 and BMP2	219
5.2.3	Chapter outline	221

5.3	Results	221
5.3.1	The structure of the BMP2-eRGMB-NEO1 _{FN56} ternary complex	221
5.3.2	The solution structure of the BMP2-eRGMB-NEO1 _{FN56} complex	230
5.3.3	BMP2 mediates clustering of NEO1 on the cell membrane	237
5.4	Discussion	244
5.4.1	RGM mediates the interaction between BMP and NEO1 signalling	244
5.4.2	BMP-induced clustering of NEO1 may be physiologically important	245
5.5	Summary	246
6	Discussion	247
6.1	The modular domain architecture of RGMs	248
6.2	Insights into Netrin-1 signalling from the RGM-NEO1 structure	251
6.3	The RGM-BMP complex may preclude BMP receptor binding	253
6.4	A model for RGM-mediated NEO1 and BMP signalling	255
6.5	Endocytosis as a signalling mechanism in other pathways	258
6.6	RGMs as targets for clinical interventions	263
6.7	Conclusions	267

Chapter 1

Introduction

1.1 Development, morphogens and cell migration

Development encompasses the transformation of a single-celled fertilised egg into a complex multicellular organism. The blueprint for this fantastically intricate process is encoded in the DNA or genome contained within that first cell. However, this does not comprise a master plan, dictating what each cell should be and its exact position within the adult. Rather, specification of a particular cell fate is imposed by interactions between cells, and the relative position of each cell within the developing organism is key for establishing which cell type it will adopt and the function it will ultimately carry out. Throughout development, cell migrations and rearrangements both precede and accompany this specification of cell type, and it is the changing interactions with different neighbouring cells that facilitate this process. Cell-to-cell communication can occur through direct contact of cells or by one cell type secreting a signal that will diffuse to another cell type and bind to receptors on that cell to trigger a signalling response. These secreted molecules, in a developmental context, are soluble proteins called morphogens. Upon receiving a specific set of signals a cell will undergo molecular differentiation, meaning it begins to synthesise a specific set of proteins and activate cellular processes that cause it to adopt a specific cell type. Specific cell types must be organised to the correct location within the developing embryo and this requires spatial determination, the position of each cell is specified by external signals before molecular differentiation occurs. Finally, following differentiation and spatial patterning, these tissues must adopt the specific shapes that they will require to carry out their functions in a process known as morphogenesis, which involves cell shape changes and cell migrations. Overall development is a complex process that requires the coordination of many different signals to direct cell movements and control cellular processes.

1.1.1 Morphogen gradients

Spatial patterning of cell fates within the developing embryo is established primarily using morphogen gradients. Morphogens are secreted signalling molecules that emanate from a specific region of tissue and diffuse away from their source resulting in a concentration gradient[1]. A morphogen acts directly on cells to produce specific cellular responses depending on its local concentration. The term morphogen, was first coined by Alan Turing in a 1952 paper outlining a theoretical model of how patterning might be produced in the early embryo using these diffusible factors[2].

*“The systems actually to be considered consist therefore of masses of tissues which are not growing, but within which certain substances are reacting chemically, and through which they are diffusing. These substances will be called **morphogens**, the word being intended to convey the idea of a form producer. It is not intended to have any very exact meaning, but is simply the kind of substance concerned in this theory.”*

Alan Turing, 1952.

The concept was refined and experimentally demonstrated in the 1960s by Lewis Wolpert using his “French Flag Model”, which described how a morphogen subdivides cells of a tissue by inducing different gene expression patterns depending on its concentration[1]. In the model this pattern is likened to the French flag: high concentrations of morphogen activate a “blue” gene, intermediate concentrations activate a “white” gene and concentrations too low to exceed the necessary threshold result in the default “red” gene expression (Fig. 1.1a). One example of this is shown for the morphogen activin, which induces different mesodermal fates: high concentrations result in activation of goosecoid expression, which specifies a dorsal fate, medium concentrations activate Xbra expression resulting in differentiation into muscle and at low activin concentrations the cells pro-

ceed along the default fate path to become vasculature (Fig. 1.1b). Much of the early work defining morphogen action and identifying morphogens was carried out in *Drosophila*. For example, mutant screens carried out by Christiane Nusslein-Volhard and Eric Wieschaus identified a plethora of morphogens involved in patterning the *Drosophila* embryo at different stages[3, 4]. Now, five major classes of morphogens have been identified, that have emerged as key regulators of embryonic development: the transforming growth factor (TGF) β (comprising activins, bone morphogenetic proteins (BMPs) and TGF β s)[5], the fibroblast growth factor (FGF)[6], the platelet derived growth factor (PDGF)[7], Hedgehog[8] and Wnt[9] families. Each family has a different patterning role in the early embryo, and although their precise roles differ across species there are some common themes; for example, BMP gradients are responsible for establishing the dorsal-ventral axis and Wnt gradients specify the anterior-posterior axis. Secreted morphogens will bind to specific receptors expressed on their target cells to activate downstream signalling pathways. The concerted action of several morphogens results in differential patterns of gene expression controlling molecular differentiation and specification of a particular cell fate within a spatial pattern.

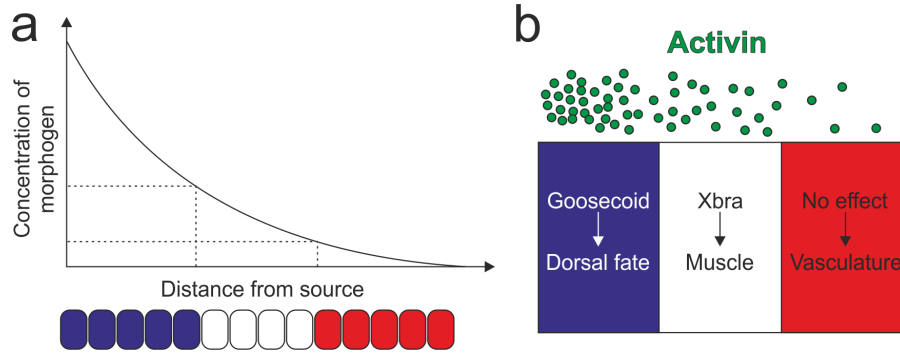


Figure 1.1: The “French Flag” model of morphogen action. (a) A graph showing the relationship between the concentration of a morphogen (y-axis) and the distance from its source (x-axis). Dotted lines indicate the threshold concentrations for inducing a particular fate on a cell. The different cell fates, corresponding to the colours of the French flag, are shown below the graph. Cells exposed to the highest morphogen concentration will adopt a “blue” fate, medium concentrations of morphogen will induce a “white” fate and below the threshold morphogen concentration cells will follow the default “red” fate. (b) An example illustrating the French Flag model. Concentration gradients of the morphogen activin control cell fates. The highest concentrations induce expression of Goosecoid (“blue”), which induces a dorsal fate; medium concentrations result in expression of Xbra (“white”), which triggers a muscle fate; and below the threshold activin concentration the default cell fate (“red”) results in a vasculature fate.

1.1.2 Overview of early development

The early stages of development transform a single cell into an embryo that is primed for further morphogenesis. The three basic tissues types are defined: ectoderm, which will go on to become epidermis and neural tissue; mesoderm, which will later be muscle, bone, connective tissue and the circulatory and lymphatic systems; and endoderm, which becomes internal organs such as the liver, lungs, stomach and intestines. The first stage of development is the cleavage phase (Fig. 1.2a,b)[10]. Following fertilisation, a series of rapid cell divisions occurs without an increase in cytoplasmic volume, these cells are called blastomeres and comprise a structure called the blastula. These repeated rounds of cell division can be observed under a microscope and occur in a very defined and synchronous manner (Fig. 1.2b). This process is initially controlled by maternal factors from the egg including the mitosis promoting factor (MPF). The blastomeres cell cycle

consists of repeated round of DNA synthesis (S-phase) followed by mitosis (M-phase), without the usual growth phases G1 and G2. After this period of rapid cell division, the control of the cell cycle by the maternal factors is rescinded and taken up by the individual cells. This is known as the mid-blastula transition, and is characterised by a loss of the synchronicity of the cell divisions. At this point, the cells begin to transcribe new mRNAs for the production of proteins that are needed for cell growth, signalling and differentiation.

“It is not birth, marriage, or death, but gastrulation which is truly the most important time in your life.”

Lewis Wolpert, 1986.

The fundamental process of gastrulation follows cleavage[10, 11]. At this stage, the blastula is converted from a simple spherical ball of cells, to an embryo with a multi-layered body plan (Fig. 1.2c). The primary germ layers, consisting of ectodermal, mesodermal and endodermal cell types, are formed and move to their correct position within the embryo. This requires the highly dynamic and concerted migration of groups of cells, which involves modulating cell motility, cell shape and cell adhesion. Cell movements and changes in cell shape result in a site of invagination called the blastopore (or primitive streak in mammals) where the mesoderm and endoderm migrate to the interior of the embryo. This destroys the blastocoel and results in the formation of a cavity called the archenteron, which will develop into the digestive tract. The embryo now exhibits the three germ layers in the correct orientation with the endoderm most central, followed by the mesoderm, and the ectoderm totally enveloping the two (Fig. 1.2c). After gastrulation, the process of neurulation results in the formation of the neural tube, which gives rise to the spinal cord and the brain. Other complex processes follow, including organogenesis and limb patterning, finally giving rise to a fully functioning organism.

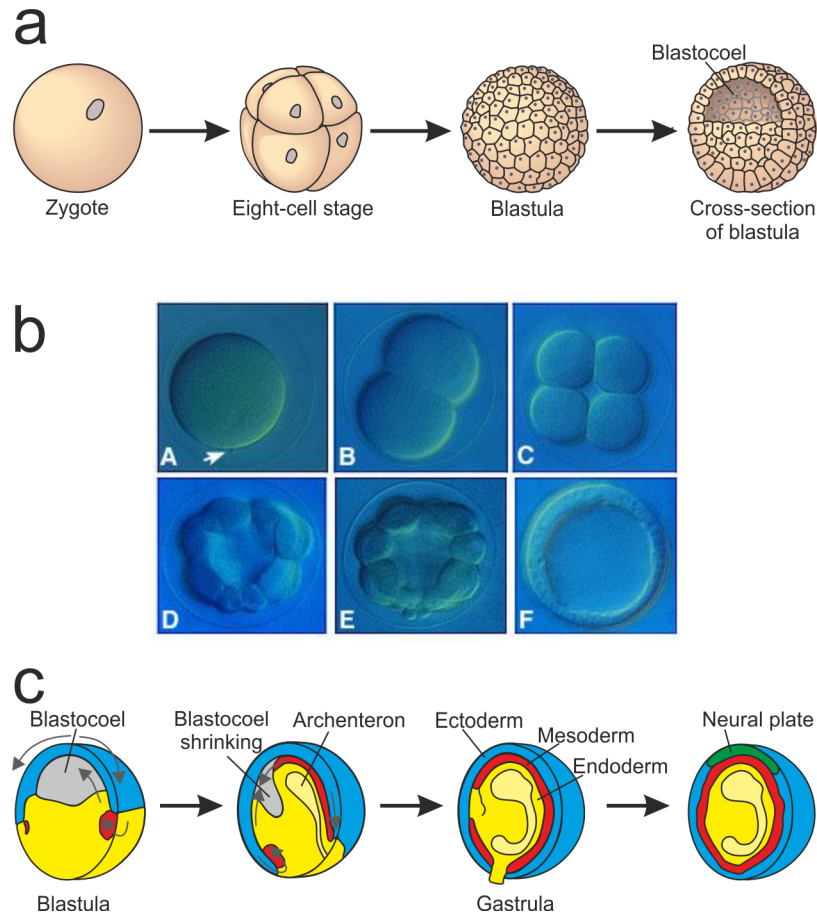


Figure 1.2: Cleavage and gastrulation. (a) Schematic showing the cleavage events in the *Xenopus* early embryo. A single cell undergoes repeated synchronous divisions without an increase in cytoplasmic volume, passing through the eight-cell stage, until a ball of cells called a blastula is formed. The blastula contains a fluid filled cavity called a blastocoel. (Image adapted from[10]). (b) Light microscopy images of sea urchin (*L. variegatus*) embryos. A = 1-cell zygote, arrow indicates site of sperm entry, B = 2-cell, C = 8-cell, D = 16-cell, E = 32-cell, F = hatched blastula. (Image taken from www.worms.zoology.wisc.edu/urchins/SUCleavage_stages.html). (c) Schematic of *Xenopus* gastrulation. Grey arrows indicate directions of cell movements. Cells invaginate at the blastopore and the endoderm and mesoderm migrate towards the interior of the embryo. The blastocoel is destroyed and a new compartment called the Archenteron is formed which will become the gut. Eventually the three germ layers are arranged surrounded by ectoderm in the form of a gastrula. A part of the ectoderm differentiates into the neural plate, which will give rise to the brain and spinal cord. Blastocoel = grey, archenteron = pale yellow, ectoderm = blue, mesoderm = red, endoderm = yellow, neural plate = green. (Image adapted from[11]).

The induction of the primary germ layers and the precise reorganisation of the embryo during gastrulation require spatial patterning even at this early stage. The anterior-posterior, dorsal-ventral and left-right axes are established through the concerted actions of many morphogens secreted largely from two organising centres. Many developmental studies into embryonic patterning and gastrulation have been carried out in model organisms such as *Xenopus laevis*, zebrafish and chick due to the ease of genetic manipulation and the availability of embryos for study. Therefore, the following examples are related to *X. laevis* development, although the general principles are directly transferable to mammalian development. The two organising centres in *X. laevis* are known as the ventral centre which secretes BMP4 and the Spemann’s organiser which is located at 180° from the ventral centre at the dorsal side (Fig. 1.3a,b). Spemann’s organiser was identified through an experiment conducted by Hans Spemann and Hilde Mangold in 1924 where they grafted a small patch of dorsal mesoderm onto the ventral side of a host embryo and found that it resulted in two-headed embryos[12] (Fig. 1.3c). They demonstrated that this “organiser” region became notochord but was capable of changing the fates of surrounding cells, and organised perfect, twin anterior-posterior and dorsal-ventral axes in the adjacent tissue[13]. Spemann’s organiser was found to secrete “dorsalising factors” which primarily comprise BMP antagonists such as chordin and noggin. These antagonists establish a concentration gradient of BMP4 activity and protect the cells at the dorsal pole from the ventralising action BMP4 (Fig. 1.3b).

The induction of a neural cell fate in the establishment of the neural plate also involves BMP4 and also Wnts and FGFs. The default fate for cells in the ectoderm is neural, but the majority of this germ layer forms epidermis as a result of the action of BMP4, which induces an epidermal cell fate. BMP antagonists are therefore important for preventing the action of BMP4, allowing cells to progress along the default path of development to form neural tissue[10]. In

chick and mouse, FGF signalling is important for specifying a neural fate by down-regulation of BMP signalling[14, 15], and Wnts appear to fulfil a similar function in *Xenopus*[16]. These signals cooperatively initiate the formation of the neural plate. During neurulation, the edges of the neural plate thicken and begin to bend upwards with the epidermis pushing in from either side to form a neural groove[11]. This eventually closes completely to form the neural tube, which will later become the brain, spinal cord and peripheral nervous system; incomplete closure of the neural tube results in neural tube defects ranging from anencephaly to spina bifida[17]. It is clear therefore, that morphogen gradients are essential for establishing the body plan during early development and in the development of the brain during neurulation.

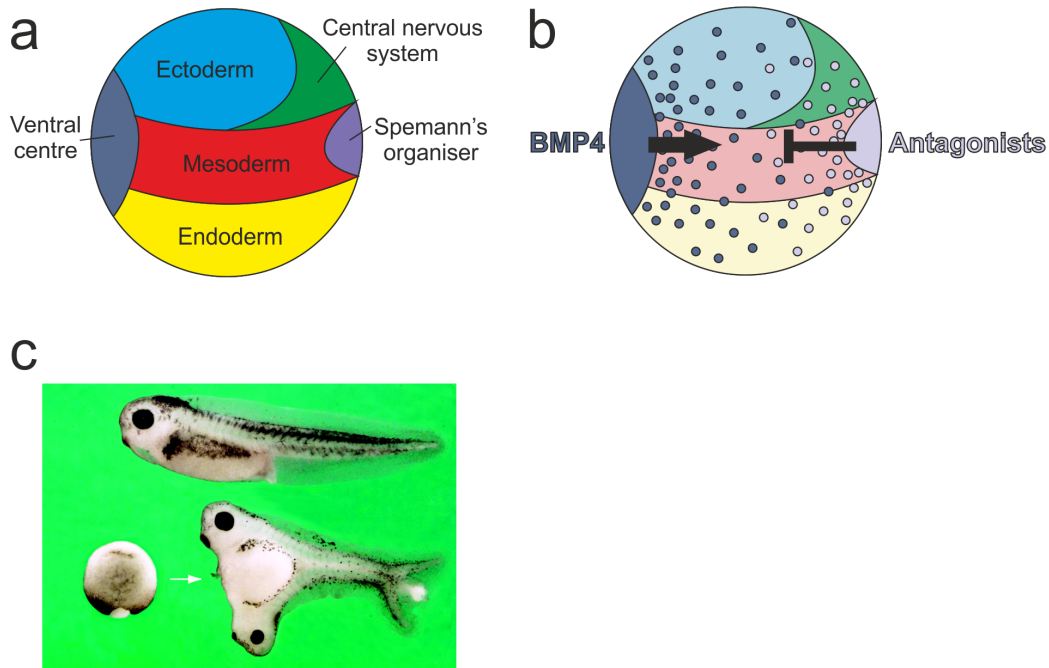


Figure 1.3: Establishing morphogen gradients in the early embryo. (a) Schematic of the *Xenopus* blastula-stage fate map showing the locations of the ventral organising centre and the Spemann's organiser. The regions that will become ectoderm (blue), neural tissue (green), mesoderm (red) and endoderm (yellow) are shown. Morphogen gradients establish this differentiation pattern. (Figure adapted from[11]). (b) One example of a morphogen gradient established by the organising centres. The ventral centre secretes BMP4 (dark purple dots), whereas Spemann's organiser secretes BMP antagonists including noggin and chordin (lilac dots) from the dorsal side. These inhibit the activity of BMP4, establishing a dorsal-ventral axis and allowing neural tissue to form. (Image adapted from[10]). (c) The Spemann-Mangold organiser experiment repeated in *Xenopus laevis*. Top = control tadpole. Bottom = tadpole after organiser graft at the same developmental stage. The graft can be visualised as the white patch on the bottom of the embryo. The graft creates a perfect, second, dorsal-ventral and anterior-posterior axis, resulting in a twin embryo. Reprinted by permission from Macmillan Publishers Ltd: [Nature Reviews Molecular Cell Biology](#) [13], copyright 2006

1.1.3 The role of cell migration in brain development

The human brain is composed of approximately 10^{11} neurons and ten times that number of supporting glial cells. An average neuron will form 10,000 connections, known as synapses, with other neural cells. The positions of all of these cells and the network of connections that they form must be precisely arranged and this relies on the ability of neurons to direct their axons over long distances, up to 2-3 feet, to establish the required connections, a process known as pathfinding. The neuronal navigator, responsible for pathfinding, is the highly dynamic, motile and sensory structure located at the tip of a growing axon, which is called the growth cone (Fig. 1.4a,b). Growth cones respond to signals in the form of molecular guidance cues, which can either attract or repel axons. These guidance molecules are either membrane tethered to a specific cell type; bound to the extracellular matrix at discrete sites; or are soluble, and diffuse from the cells that secrete them, through the extracellular environment, establishing a concentration gradient. Guidance cues bind to target receptors in the membrane of axon growth cones to activate downstream signalling pathways that result in changes in the organisation of the cytoskeleton thereby triggering either growth cone expansion or collapse, which ultimately drives axon growth towards or away from the signal respectively[18, 19].

The overall architecture of the growth cone is established by two elements of the cytoskeleton, actin microfilaments and microtubules (Fig. 1.4b,c). Growth cone motility is associated with projections of the membrane in the form of filopodia (thin, cylindrical tendrils) and lamellopodia (broad, flat sheets), which are brought about largely by actin polymerisation (Fig. 1.4c). Growing actin filaments push the membrane forward, over the surface that the growth cone is migrating upon, and the new structures are anchored by transmembrane proteins such as integrins[20]. Microtubule bundles provide structural support to the axon

and the base of the growth cone (Fig. 1.4c). Some overlap between actin filaments and dynamic microtubules in the regions near filopodia occurs and this is thought to be involved in stabilising filopodia in response to attractive guidance cues[20, 21, 22]. The growth cone also contains the motor protein myosin, which can bind to actin and exert a motile force relative to either other actin filaments or to the cell surroundings. Myosin appears to play a role in the retrograde flow of actin (away from the filopodia tips)[23, 24, 25] and also in the growth cone collapse that is induced by repulsive guidance cues[26, 27, 28]. Attractive guidance cues cause growth cone expansion and this directs axon growth towards their source, whereas repulsive guidance cues trigger growth cone collapse causing axons to avoid them (Fig. 1.4d,e). Attractive guidance cues activate the small guanosine-5'-triphosphate hydrolysing enzymes (GTPases) Rac and Cdc42 that enhance actin polymerisation and so drive growth cone expansion and subsequent axon migration[29, 30, 31, 32] (Fig. 1.4d,e). Repulsive guidance cues activate small GTPases of the Rho family that activate Rho-kinase (ROCK), which phosphorylates and activates myosin to trigger collapse of the growth cone[26] (Fig. 1.4b).

There are several families of guidance cues including semaphorins[33], ephrins[34], netrins[35] and Repulsive Guidance Molecules (RGMs)[36]. Guidance cues are categorised as having either repulsive or attractive effects and mediate these by binding to cell-surface receptors. However, often multiple cues utilise the same cell-surface receptor but trigger different downstream effects, for example both netrin-1 and RGM bind to the same receptor, Neogenin (NEO1) but netrin-1 will trigger an attractive cue whereas RGM will signal a repulsive one[37]. Therefore, the paths different axons take through the brain will depend on the axon guidance cues available in the extracellular surroundings and on the receptors that the growth cone of each axon is expressing. Interestingly, recently, morphogen proteins such as those commonly associated with embryonic patterning

including Wnts[38] and BMPs[39], are being increasingly implicated in having roles in axon guidance. This is perhaps not surprising since morphogens and guidance molecules share many overlapping characteristics, for example they are both secreted to form gradients and trigger multiple graded responses in their target cells; morphogens are important in regulating cell motility and adhesion during development.

We have discussed the essential roles of cell migration in axon guidance, gastrulation, the development of the embryo and morphogenesis; however, the roles of cell migration are not limited to these functions. In related systems, migrating fibroblasts and epithelial cells respond to wounds and direct tissue regeneration, immune cells patrolling the blood stream migrate into tissues to deal with infections, and in cancer metastasis tumour cells migrate into the blood stream and then invade new environments. The same basic migration and guidance processes as those described above for axons are utilised, however the migration occurs in different tissue contexts, using different molecular repertoires and extracellular guidance cues[40]. Although, as we will see below, it is possible for one type of guidance cue to be involved in many of these cell migration systems.

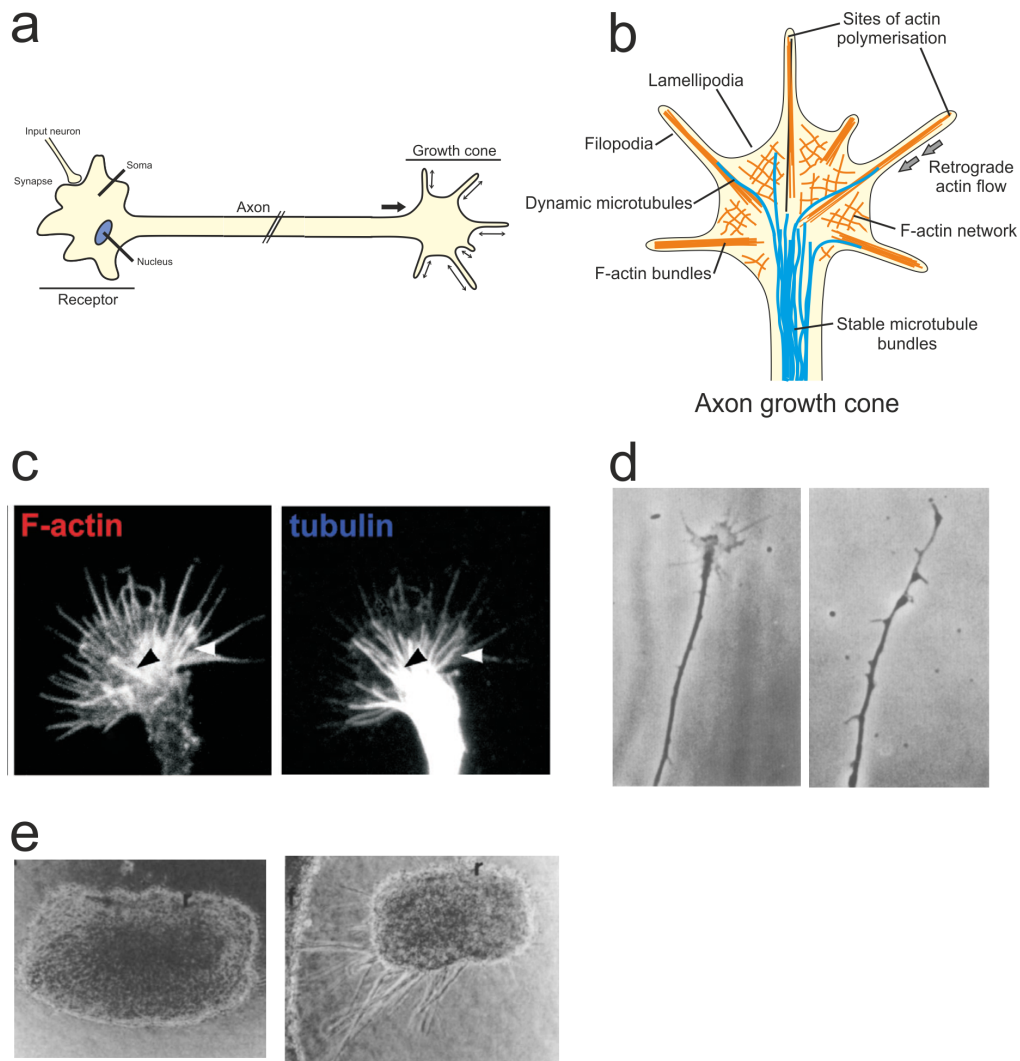


Figure 1.4: Axon growth cones: structure and guidance. (a) A schematic showing the features (labelled) of a neuronal cell. Axon of variable length indicated by '//'. Arrows indicate directions of growth (Figure adapted from[10]). (b) Schematic showing the architecture of the growth cone. F-actin polymers (orange), microtubules (blue) and retrograde actin flow (purple arrows) are shown. (Figure adapted from[41]). (c) Immunofluorescence confocal images of fixed growth cones from embryonic cortical neuron cultures. Actin filaments (F-actin, stained with phalloidin dye) and microtubules (tubulin, labelled with an antibody) are shown. Arrowheads indicate points at which the two co-localise. (Image from[41]). (d) Light microscopy images of dorsal root ganglion cell cultures showing an expanding growth cone (left panel) and a growth cone after collapse (right panel). Collapse induced by cell medium containing collapsin. (Image reprinted from [33], Copyright 1993, with permission from Elsevier: Cell). (e) Light microscopy images of the growth of commissural axons over 40 hours from dorsal spinal cord explants from rat embryos embedded in a collagen matrix. Without a guidance cue, no axon outgrowth observed (left panel), a floor plate explant secreting an attractive guidance cue, resulted in directed axon growth (right panel). (Image from[35]).

1.2 The discovery of Repulsive Guidance

Molecules

Investigations into the question of how the precise mapping of the central nervous system (CNS) is accomplished began using the chick retinotectal model system. In the chick embryo approximately 2 million neurons from each eye invade the target organ, the optic tectum, and form synapses with their target neurons[42]. Retinal ganglion cells project axons into the optic tectum in a topographic manner; temporal retinal axons project to the anterior side of the tectum, whereas nasal retinal axons invade the posterior side. The theory that gradients of guidance cues might direct this process sparked a series of *in vitro* experiments in search of these directional cues. Two assays called the stripe and collapse assays proved to be important tools in these investigations[43, 44]. In the stripe assays alternating parallel stripes of membrane from the anterior and posterior tectum were arranged and showed that temporal axons avoided the posterior lanes, whereas nasal axons showed no such preference[45]. Similarly, in the collapse assay, vesicles composed of posterior membrane triggered growth cone collapse in temporal but not nasal axons[43]. However, although the posterior membranes were shown to have these growth cone collapse and axon guidance properties, it still remained to identify the active molecular guidance cue.

Repulsive Guidance Molecule (RGM) was the first candidate for this guidance cue, and was identified in 1990. It was originally described as a 33 kDa membrane-bound glycoprotein, and vesicles enriched with this protein were highly active in collapse and stripe assays[46]. The protein was found to be sensitive to cleavage by phosphatidylinositol-specific phospholipase C (PI-PLC) indicating that it was tethered to the membrane by a glycosylphosphatidylinositol (GPI)-anchor[46, 47]. Subsequently, a monoclonal antibody (F3D4) raised against this protein was used to show that the protein was expressed in a gradient from high

in the posterior to low in the anterior of the tectum[48] (Fig. 1.5a). Furthermore, when used in conjunction with a technique called chromophore-assisted laser inactivation that accurately targets proteins bound by an antibody for inactivation, the F3D4 antibody specifically blocked the repulsive guidance activity of the protein. It was on the outcome of these experiments that this protein was named RGM.

It was not until 2002 that the gene encoding chicken RGM (cRGM) was identified and a full-sequence analysis could be carried out[36]. Recombinant cRGM showed a graded expression pattern from low in the anterior to high in the posterior of the chick tectum and the isolated protein demonstrated full repulsive guidance activity in stripe and collapse assays, similar to that induced by posterior tectum membranes[36] (Fig. 1.5b,c). This verified that the correct gene had been identified. Chick RGM is a 432 amino acid glycoprotein and shows no sequence homology to other guidance cues. The protein comprises an N-terminal signal peptide, an RGD site thought to be involved in cell adhesion[49], a partial von Willebrand factor type D (vWfD) domain[50] and a GPI-anchor[36] (Fig. 1.6a). Interestingly, the calculated molecular weight of the protein (49 kDa) was higher than that reported experimentally for RGM (33 kDa). Peptide sequencing to determine the amino-terminus of the active chick RGM resulted in the sequence "PHLRT", which corresponded to the lower molecular weight form of the protein, and it was suggested that the N-terminal 149 amino acids are cleaved off by an unknown protease[36].

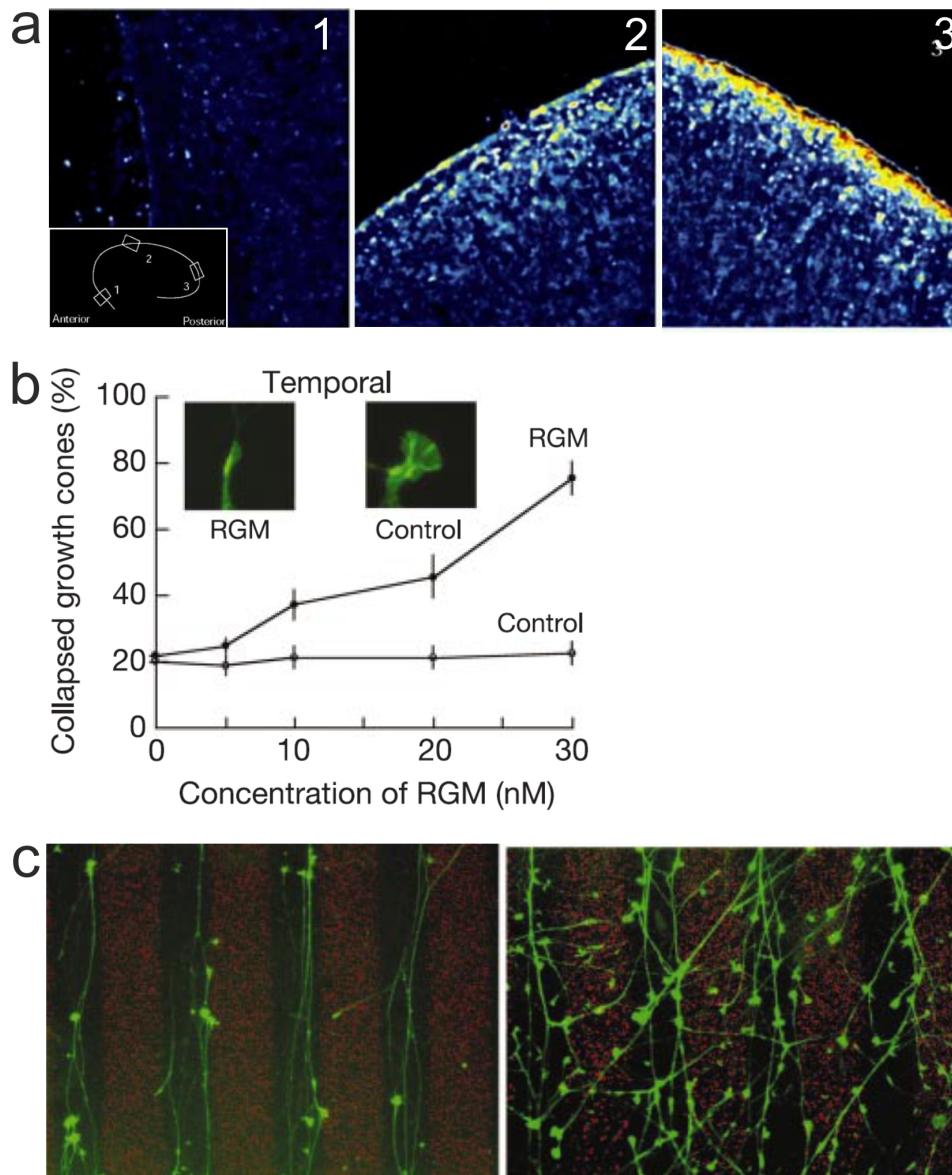


Figure 1.5: The discovery of RGMs. (a) Pseudo-coloured photographs of sections from the anterior (1), medial (2) and posterior (3) parts of the chick optic tectum after immunostaining with F3D4 anti-cRGM antibody. F3D4 immunoreactivity increases from the anterior to the posterior indicating an anterior-posterior gradient of cRGM distribution. (Image reprinted from [Current Biology](#) [48], Copyright 1996, with permission from Elsevier). (b) Graph showing the percentage of collapsed growth cones from temporal neurons in the presence of membranes from RGM- or control-transfected HEK 293 cells. RGM effectively induces growth cone collapse. (Image reprinted by permission from Macmillan Publishers Ltd: [Nature](#) [36], copyright 2002). (c) A stripe assay showing that temporal retinal axons avoid stripes derived from membranes of RGM-transfected HEK 293 cells (labelled with red fluorescent beads) and grow on mock-transfected cell membranes (left panel). Nasal retinal axons, however, do not show this preference (right panel). (Image reprinted by permission from Macmillan Publishers Ltd: [Nature](#) [36], copyright 2002).

There are orthologues of chRGM found in all vertebrates, with varying levels of similarity, from human with 82 % amino acid identity, to zebrafish with 69 % amino acid identity (Fig. 1.6b). Orthologues are also found in invertebrates, although, as expected these show a higher sequence divergence, for example the DRAG-1 protein in the roundworm *Caenorhabditis elegans*, with 33 % amino acid identity. Furthermore, following on from the discovery of chicken RGM, three mammalian orthologues were identified by sequence analysis and named RGMA, RGMB (also called DRAGON) and RGMC (also called hemojuvelin (HJV), HFE2, DRAGON-like muscle (DL-M))[51, 52] (Fig. 1.6b). Of these RGMA shares the closest sequence homology with chRGM (78 % amino acid identity). Human RGMA shares 50 % amino acid identity with RGMB and just under 47 % with RGMC (Fig. 1.6b). Concurrent with their discovery by sequence analysis, RGMB and RGMC were also identified by biochemical methods, hence why they have alternative names. The first of these studies used a genomic binding site strategy to identify targets of DRG11, a transcription factor expressed in dorsal root ganglion (DRG) sensory neurons[53]. One protein up-regulated by DRG11 was named DRAGON, reflecting that it was turned “on” in the DRG. Mouse homologues to DRAGON were also identified from mouse genomic databases, yielding mRGM (i.e. RGMA) and a protein named both for its similarity to DRAGON and its expression pattern in the mouse embryo: DRAGON-like muscle (i.e. RGMC)[53]. In another study, mutations associated with a subtype of an early-onset iron overload disease called juvenile hemochromatosis (JHH) were mapped to the human gene HFE2, and the protein product of this gene was named hemojuvelin, which corresponds to RGMC[54]. For clarity, in this thesis RGM proteins will be referred to as RGMA, B and C.

The pattern of tissue expression for RGMs is wide-ranging and distinct for each family member[51, 55, 52]. *In situ* hybridisation studies during the early development of mouse embryos and in brain tissue sections, using antibodies against

the three RGM family members, showed that the most abundant expression of RGMA and RGMB is found in the nervous system where they occupy non-overlapping but complementary regions[51, 52]. Outside of the brain, RGMA and RGMB expression has been found in the heart, lung, liver, bone, skin, kidney, the reproductive system and gut, hinting at further diverse functions[56, 55, 57]. In contrast, RGMC is exclusively expressed in all skeletal and cardiac muscle and not in the central nervous system[55].

The RGM family shares common structural features, many of which were identified in cRGM, including an N-terminal signal peptide and a putative RGD motif (conserved in RGMA and RGMC), which is thought to be important for integrin-mediated adhesive function[58, 36, 59]. They also have a partial vWfD domain, comprising about 40 % of a full-length domain that contains the highly conserved GDPH cleavage site, which was initially suggested in cRGM to be cleaved by an unknown protease. However, studies using recombinant RGMC suggest that RGMs possess autocatalytic activity and show that the aspartic acid-proline bond of the cleavage site is acid-labile and incubation of RGMC at pH 5.5 enhances autocatalytic cleavage[42]. The remaining portion of the RGMs is composed of a hydrophobic domain of unknown structure followed by a GPI-anchor, although RGMs can be cleaved by proteases and phospholipase C to remove this anchor and produce a soluble form allowing for long-range actions[36].

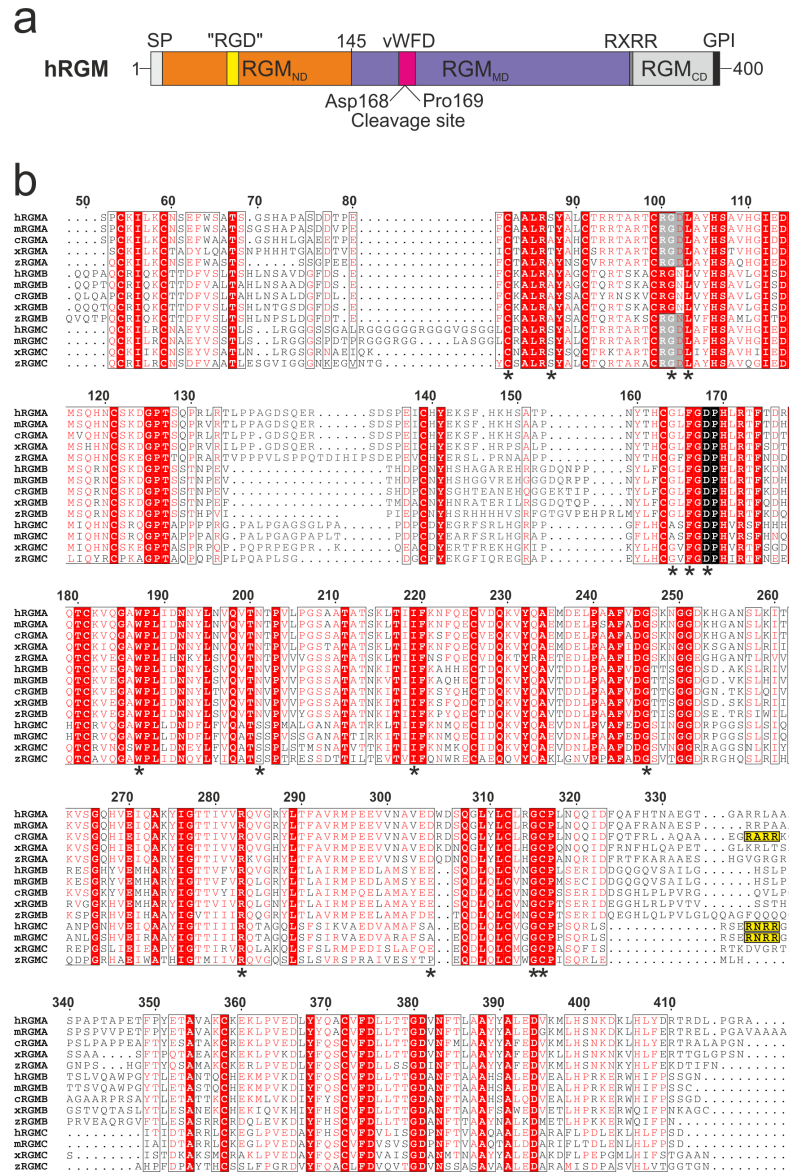


Figure 1.6: Domain organisation and sequence alignments of the vertebrate RGM family members. From [240], [Science](#), reprinted with permission from AAAS. (a) Schematic of the domain organisation of human (h) RGM. Numbering for hRGMC. N-terminal domain (RGM_{ND}) (orange), RGD motif (yellow), middle domain (RGM_{MD}) (purple), partial von Willibrand factor type D domain (vWFD) (pink), Arg168-Pro169 autocatalytic cleavage site and C-terminal domain (RGM_{CD}) (grey) are highlighted. GPI = glycosphosphatidylinositol anchor (black), SP = signal peptide, RXRR = putative cleavage site. (b) Alignment of human (h), mouse (m), chicken (c), *Xenopus* (x) and zebrafish (z) RGM A, B and C sequences. Aligned using MULTALIN (bioinfo.genotoul.fr/multalin/multalin.html) and formatted with ESPRIPT (esprpt.ibcp.fr/ESPrpt/ESPrpt/). Numbering corresponds to hRGMB (including secretion signal). RGD motif (grey box), autocatalytic cleavage site (black) and putative furin cleavage site (yellow box) are highlighted. Juvenile hemochromatosis-linked mutations in hRGMC are marked with asterisks (*).

1.3 Functions of RGMs

1.3.1 RGMA

Given the function of chicken RGM as a repulsive axon guidance cue in the retinotectal system[60, 36], the mammalian orthologue, RGMA, was tested for a similar function[60, 51]. However, in contrast to chRGM, no gradient in mouse RGMA expression was detected along the anteroposterior axis of the superior colliculus, which is the mammalian equivalent of the optic tectum; although a clear gradient was shown for another repulsive guidance cue, EphrinA5. Furthermore, in an RGMA^{-/-} knockout mouse no defects in neuronal patterning were observed and the loss of RGMA did not seem to impair anteroposterior mapping of retinal axons. However, approximately 50 % of RGMA^{-/-} embryos showed incomplete closure of the neural tube resulting in an exencephalic phenotype, where brain structures are exposed to the external environment[51]. Other studies have also identified a role for RGMA in early neural development, for example, work in *Xenopus* embryos has shown that knockdown of RGMA, performed by injection of morpholinos, results in severe morphological defects of the eye and affects the expression of marker genes of the developing forebrain[61]. In addition, a role for RGMA has been suggested in the early development of the mouse intestine due to the expression of RGMA in the intestinal ganglia and epithelia at specific developmental stages[56]. RGMA may therefore play a role in vertebrate early neural and organ development that is quite distinct from its roles in axon path finding in the more mature nervous system.

Despite this lack of evidence for RGMA as a repulsive axon guidance cue in mammalian retina development, many studies have demonstrated a role for RGMA in other axon guidance systems. For example, RGMA was shown to be involved in establishing the specific laminar pattern of afferent neuron terminations in the de-

veloping mouse hippocampus. RGMA repelled entorhinal axons in stripe assays; and in entorhino-hippocampal co-cultures neutralising antibodies against RGMA resulted in entorhinal axons entering inappropriate hippocampal areas[62]. Subsequently, work in the developing *Xenopus* forebrain, showed that a knockdown of RGMA using morpholinos resulted in patterning defects, specifically a failure to establish the supraoptic tract[37]. This was the first *in vivo* evidence, outside of the chick model system, of the function of RGMA in axon guidance. Data supporting this role for RGMA were also generated by a study showing that injection of RGMA mRNA into developing *Xenopus* embryos, causing over-expression of RGMA, resulted in distorted pathfinding[63]. Interestingly this treatment also triggered sprouting of ectopic neurons, showing not only a role for RGMA in axon guidance, but potentially also in neuronal cell differentiation. In addition to these *in vivo* studies, RGMA has also been shown to inhibit neurite outgrowth of several different types of neurons using classical, *in vitro*, assays for growth cone collapse and axon outgrowth[64, 65, 66]. Further studies *in vivo* are required to determine the precise function of RGMA in axon pathfinding, particularly in mammalian systems. It may be that a double knockout of $\text{RGMA}^{-/-}$ and $\text{RGMB}^{-/-}$ is required to observe a phenotype as it has been suggested that RGMB might be able to compensate for loss of RGMA in some systems[51]. However, the induction of growth cone collapse to inhibit the outgrowth of neurons appears to be a general function of RGMA.

Related to this, RGMA has been implicated as a major inhibitor of axon regeneration after injury in the mammalian central nervous system[67, 68, 69]. Following spinal cord injury in rats, RGMA expression increases around the lesion site[67, 68]. A similar expression pattern was observed for RGMA in adult human brains following traumatic brain injury or focal cerebral ischemia, with expression levels maximal 24 hours post-trauma and elevated for weeks into months afterwards[69]. The presence of RGMA at the lesion site, combined with

its known ability to inhibit axon growth suggests that it prevents nerve regeneration in the scar tissue. Indeed, administration of an activity-blocking anti-RGMA antibody promoted new growth of corticospinal tract axons and resulted in enhanced motility recovery in rats after spinal cord injury[67]. Likewise, RGMA was also shown to have a similar inhibitory affect on regeneration of damaged auditory neuron fibres in the mouse inner ear[70]. Once again, anti-RGMA antibodies disrupted this inhibition, promoting enhanced growth of neuron fibres and increasing levels of synaptogenesis with their target cells. These studies show that RGMA is an inhibitor of neuronal regeneration at sites of injury, both within the CNS and in the peripheral nervous system (PNS). This function is most likely linked to RGMA-induced growth cone collapse, which inhibits neurite outgrowth.

Directed cell migration is not only required in development and regeneration of the nervous system, it is also essential for appropriate and localised immune responses. The migration of leukocytes into inflamed tissue shares the basic mechanism of cytoskeleton remodelling with axon guidance[71]; and has also been shown to be under the control of attractive and repulsive guidance cues, some of which have parallel functions in neuronal path finding[72, 73, 74, 75, 76, 77]. RGMA is strongly expressed in epithelium, leukocytes and highly vascularised tissues such as lungs and kidneys that are sensitive to infection, hinting at a role in leukocyte chemotaxis[78]. Recently, it was shown that RGMA inhibits leukocyte migration and attenuates the inflammatory response. Zymosan-A, an experimental inducer of inflammation, was used to induce peritonitis (inflammation of the lining of the abdomen) in mice. Injection of RGMA in this system acted as an inhibitor of leukocyte recruitment and diminished the signs of acute inflammation[78]. Furthermore, RGMA was shown to play a role in the activation of CD4+ T-cells, which is the primary trigger of an inflammatory response that results in axonal damage contributing to the development of the symptoms of

multiple sclerosis (MS)[79]. Antibodies against RGMA were used to demonstrate that RGMA controls activation of T-cells and the induction of various proteins comprising the immune response, including interferons and interleukins. These studies characterise RGMA as a potent immuno-modulatory protein.

RGMA also appears to be involved in apoptosis in developing embryos, across many vertebrates. For example, RGMA has been identified as a cell survival factor in the developing chick neural tube and in the mouse retina. Over-expression of the RGM receptor NEO1, or down regulation of RGMA, induced apoptosis; and addition of RGMA increased cell survival[80, 60]. This identified NEO1 as a dependence receptor and identified RGMA as a controller of cell survival, although the precise mechanism for this process remains controversial[81]. It has also emerged that the activity of RGMA contributes to neurodegeneration in a mouse model of multiple sclerosis (Experimental Autoimmune Encephalomyelitis (EAE)) and that neutralising antibodies against RGMA reduce neuronal cell death in this model without altering inflammatory responses[82]. RGMA has been shown therefore to be a controller of cell survival in two different systems.

Finally, RGMA was recently identified as a co-receptor for BMP signalling, a pathway best known for its role in development[83]. In a luciferase reporter assay transfection of RGMA cDNA into cells enhanced BMP but not TGF β signalling, and RGMA was shown to bind directly to BMP ligands and also to the BMP type I receptor. In rat spinal cord RGMA is expressed in neurons that also show active BMP signalling suggesting that RGMA functions as a co-receptor for this morphogenetic pathway. However, the specific physiological role of RGMA-activated BMP signalling has not yet been determined.

In summary then, the functions of RGMA are wide-ranging but can be grouped into five categories: axon repulsion, neural development, immune cell regulation, cell survival and activation of BMP signalling.

1.3.2 RGMB

The expression pattern of RGMB is in general very similar to that of RGMA, with strong expression observed throughout the nervous system. It is therefore reasonable to expect RGMB to be involved in axon guidance, similar to chRGM and RGMA. However, in an RGMB^{-/-} knockout mouse no neuronal patterning or developmental defects were observed[84]. Although, RGMB has been shown to inhibit neurite outgrowth in mouse cerebellar granule neurons, at least *in vitro*[85]. Furthermore, in the developing mouse dentate gyrus RGMB was shown to abolish migration of neuroepithelial cells expressing the receptor NEO1 both *in vitro* and *in vivo*, suggesting that RGMB plays a role in neuronal cell migration[86]. In a role that may be related to axon pathfinding, transfection of HEK293T cells with RGMB enhanced the adhesion of cultured DRG neurons to these cells suggesting that RGMB has adhesive properties[53]. Immunoprecipitation experiments suggested that homophilic interactions between RGMB molecules might be responsible for the observed RGMB-mediated cell-cell adhesion. Several other cell-surface GPI-anchored proteins, including NCAM (neuronal cell adhesion molecule), ephrins and tenascin, mediate adhesion of neurons to non-neuronal cells or components of the extracellular matrix. Interestingly, this was in contrast to RGMA, which in the same experiment resulted in a two-fold decrease in adhesion of DRG neurons to transfected HEK293T cells[53]. Effective adhesion of cells to their surroundings is essential for migration. The complementary expression pattern for RGMA and RGMB combined with their different effects on DRG neuron adhesion may have consequences for axon guidance within the CNS.

Overall the evidence for RGMB acting as an axon guidance molecule *in vivo* is limited; nevertheless RGMB has many other functions in common with RGMA. First, RGMB is expressed in neurons, oligodendrocytes and myelin in the adult

rat CNS and is upregulated at sites of spinal cord injury[85]. This combined with the neurite outgrowth inhibitory activity of RGMB suggests that it too may function as an inhibitor of axon regeneration in the CNS. Secondly, RGMB has a role in apoptosis and acts as a controller of cell survival, like RGMA, although this function was identified in renal tubular epithelium rather than neuronal cells. RGMB over-expression increased renal epithelial cell death, and knockdown of RGMB in RGMB^{+/-} mice showed a decrease in kidney cell apoptosis compared to controls, under hypoxic conditions[87]. This is in contrast to observations where over-expression of RGMA increased cell survival, another example of the contradictory effects of these two proteins. Interestingly, in this same study RGMB was shown to inhibit expression of E-cadherin a cell-cell adhesion molecule[87], a function that was not demonstrated for RGMA. Finally, RGMB has also been implicated in the immune response. RGMB is highly expressed in macrophages and acts to inhibit the expression of interleukin-6 (IL-6), a pro-inflammatory cytokine[88]. Therefore both RGMA and RGMB attenuate the inflammatory response, although they appear to do so via different mechanisms.

Interestingly, the inhibition of the expression of IL-6 by RGMB was found to be dependent on levels of BMP signalling. RGMB was the first RGM family member to be identified as a co-receptor for BMP signalling[89] and many of its physiological roles have subsequently been shown to be dependent on this signalling pathway. The expression pattern of RGMB in the developing *Xenopus* embryo is overlapping with that of the BMP receptors and using a luciferase reporter assay RGMB was shown to activate BMP signalling and to enhance the sensitivity of cells to BMP ligands *in vitro*[89]. RGMB also enhances the sensitivity of cells to BMP signalling in the *Xenopus* embryo, lowering the threshold required for downstream gene activation, and this appears to be important for patterning of the developing nervous system[89]. Another potential function of RGMB signalling through BMPs is the regulation of tight junctions in the epithelia of

adult kidney tubules. Tight junctions are important for maintaining the polarity of the cells and are essential to control transport of small molecules across the epithelial barrier; transepithelial resistance (TER) can be used to measure their effectiveness and BMP signalling has been shown to enhance TER[90, 91]. Both RGMB and BMP4 are expressed in kidney epithelial cells and co-transfection of cells from a mouse inner medullary collecting duct cell line (mIMCD3) with RGMB and BMP4 enhanced TER over that stimulated by BMP4 alone, suggesting that RGMB enhances BMP4 signalling to increase the stability of tight junctions in the kidney[92]. Despite several studies into this area, the role of RGMB-enhanced BMP signalling has not yet been fully elucidated, and in fact the physiological role of RGMB itself is still not clear although it appears to be a protein with many different functions.

In summary, RGMB appears to be important in cell adhesion, immune cell regulation, axon growth inhibition and cell survival; functions it largely shares with RGMA, although the two often have very different roles within these systems. RGMB is an important activator of BMP signalling and this appears to be important in brain and kidney development. Overall, the precise biological functions of RGMB are not well defined.

1.3.3 RGMC

The primary function of RGMC is in the regulation of iron homeostasis. Iron is an essential co-factor for many physiological processes including oxygen transport and respiration, but free iron is also toxic as its electron transfer properties mean it can result in the creation of harmful free radicals; and so the body tightly regulates circulating iron levels[93]. Iron is stored both in macrophages that recycle iron resulting from the destruction of red blood cells, and the liver. Iron export from these stores is mediated by ferroportin-1, the only known transporter of

iron efflux from cells[94, 95, 96]. This transporter is regulated by the peptide hormone hepcidin, which is secreted by hepatocytes and binds to and directs the internalisation of ferroportin-1 for degradation[97]. Removal of ferroportin-1 prevents iron export, causing a reduction in blood iron levels. Mutations in the gene for hepcidin result in iron overload diseases[98, 99], whereas increased expression of hepcidin causes iron deficiency and anaemia[100, 101]. Expression of hepcidin is activated by BMP6 signalling through the downstream SMAD pathway, as demonstrated by the massive iron overload exhibited in a liver specific SMAD4^{-/-} knockout mouse[102, 103, 104]. Additionally, under inflammatory conditions hepcidin expression can also be activated by interleukins through the downstream JAK/STAT signalling pathway[105].

RGMC was initially discovered when mutations in the gene were linked to the blood iron overload disorder juvenile hemochromatosis (JHH), which provided the first evidence that RGMC was involved in the process of iron regulation[54]. RGMC^{-/-} knockout mice exhibit high levels of iron accumulation in the liver, pancreas and heart, and show a decrease in hepcidin expression coupled with a corresponding increase in ferroportin levels; suggesting that RGMC regulates levels of hepcidin to control iron storage[106]. RGMC, like the other RGMs was subsequently identified as a co-receptor for BMP signalling using a luciferase reporter assay[102]. Furthermore, induction of hepcidin expression by BMP was enhanced by transfection of HepB3 liver cells with RGMC and diminished in RGMC^{-/-} hepatocytes, demonstrating that activation of BMP signalling is the mechanism by which RGMC controls hepcidin expression. Experiments showed that BMP6 is the endogenous regulator of hepcidin expression as BMP6^{-/-} mice have a phenotype resembling that of RGMC^{-/-} mice, exhibiting reduced hepcidin expression and tissue iron overload[107, 108]. However, activation of BMP6 signalling by RGMC is essential for maintenance of normal serum iron levels as in RGMC^{-/-} knockout mice hepcidin expression is low, despite elevated BMP6

levels[109]. It is clear therefore, that RGMC signalling through BMP6 plays a major role in the regulation of normal blood iron levels.

It appears that membrane attachment of RGMC is essential for this function and in fact soluble RGMC can bind to BMP and act as an antagonist for membrane-bound RGMC, resulting in reduced hepcidin expression[110, 111]. There is a circulating pool of soluble RGMC in human and mouse serum of around 1 $\mu\text{g}/\text{ml}$ and 0.58 $\mu\text{g}/\text{ml}$, respectively and the ratio between soluble and membrane-bound RGMC may therefore exert fine control over serum iron levels[112, 113]. However, whilst injection of Fc-tagged RGMC into mice inhibits hepatic BMP signalling resulting in reduced hepcidin expression and increased serum iron levels[103], injection of anti-RGMC antibodies into rats and monkeys also down-regulates hepcidin expression and increases serum iron levels[114]. Furthermore, patients with iron-deficiency anaemia were found to have lower circulating soluble RGMC levels than those in healthy patients[112]. It is therefore not yet clear what role soluble RGMC plays *in vivo*.

1.4 The role of RGMs in disease

1.4.1 Cancer

Both RGMA and RGMB have been implicated in cancer development and progression. This is perhaps unsurprising given their role as important co-receptors for the BMP signalling pathway, which has well-established links to cancer progression, particularly in determining the aggressiveness of primary tumours and in tumour metastasis (reviewed in[115]). Additionally, the RGM receptor NEO1 is a close homologue of deleted in colorectal cancer (DCC), which plays critical roles in colorectal cancer (CRC)[116, 117, 118, 119]. The involvement of other

axon guidance cues in cancer progression is an emerging phenomenon[120, 121, 122, 123, 124, 125], these cues regulate angiogenesis, cell survival, cell migration and cell adhesion, all essential processes in cancer development and all processes in which RGMs have been implicated. It was shown that RGMA expression is downregulated across colon cancer cell lines, CRCs and adenomas, suggesting that RGMA inactivation is a crucial step in CRC development[126, 127]. The mechanism for this downregulation appears to be methylation of the gene promoter: the average methylation level in normal colon is below 10 %, but in CRCs the level of methylation was found to be over 85 %[126, 127]. RGMA appears to suppress the migration and invasion abilities of CRCs and promote apoptosis upon DNA damage in in vitro assays, and given its inactivation in CRCs, it appears to function as a tumour suppressor[126]. In contrast, a study into the role of RGMB in CRCs showed that RGMB expression was upregulated in CRC tissues and that RGMB promoted the cell proliferation, migration and invasion capacities of CRC cell lines[128]. Furthermore, in mouse CRC xenografts, a RGMB knockdown resulted in significantly slower tumour growth, identifying RGMB as a target for potential therapies to treat colon cancer[128]. However, data collected for RGMB function in breast cancer development contradicts this. A knockdown of RGMB in two different breast cancer cell lines resulted in an increase in cell proliferation, migration and adhesion, and a decrease in levels of apoptosis[129]. Therefore, in this disease model, RGMB seems to function as a tumour suppressor, similarly to RGMA. It may be that RGMs have very different roles in different tissue types, which may depend on the relative activities of the two pathways that they signal through: BMP and NEO1 signalling systems.

1.4.2 Multiple Sclerosis

The primary trigger of multiple sclerosis in humans is the immune response initiated by the activation of T-cells, which are located in the myelin sheath of neurons in the brain and spinal cord. This results in demyelination, neuronal degeneration and the progressive paralysis that is associated with MS. Activation of T-cells requires sustained interactions between T-cell receptors (TCRs) and their corresponding antigens on the surface of antigen presenting cells such as dendritic cells. RGMA is a potent immunomodulatory protein with roles in the inflammatory response and notably, mutations in RGMA are associated with multiple sclerosis (MS) and correlate with changes in interferon levels in the cerebrospinal fluid of patients[130]. Recently the role of RGMA in MS was investigated using a mouse model of MS called EAE[79].

It was shown that the expression of RGMA is upregulated in the dendritic cells of mice after induction of EAE, and that RGMA is expressed in dendritic cells from brain and spinal cord samples from patients with MS, but not control samples. The RGMA receptor NEO1 is expressed in both mouse T-cells and sections from the brain and spinal cord of MS patients, consistent with RGMA playing a role in the activation of T-cells. Importantly, administration of RGMA-specific antibodies to mice with EAE reduced both the severity of the disease and the percentage of mice presenting with symptoms. These mice showed a significant reduction in both myelin loss and axonal damage. In cell samples from MS patients, anti-RGMA antibodies reduced a subset of types of T-cell activation and down-regulated the expression of proteins of the immune response including interleukins 2, 4 and 17 and interferon- γ . This and other studies have identified RGMA as a promising molecular target in the treatment of MS and highlighted the clinical potential of RGMA-specific antibodies[79, 82].

1.4.3 Regenerative failure

Unlike those of the PNS, axons of the CNS demonstrate no functional regeneration after injury, which is generally the cause of paralysis after a CNS injury. The local glial cell environment of the axons has been implicated as the major cause of this lack of regeneration. This encompasses inhibitor proteins including neurite outgrowth inhibitor A (Nogo-A), myelin-associated glycoprotein (MAG), oligodendrocyte-myelin glycoprotein (OMgp) and most recently RGMs. RGMs are upregulated at sites of injury in rats[68] and after focal cerebral ischemia and traumatic brain injury in humans[69]. Importantly, antibodies against RGMA administered at the site of spinal cord injury in rats resulted in a significant improvement in motility, which appears to be due to increased axon sprouting and growth of neuron fibres at the lesion site[67]. This suggests that antibodies targeted to RGMA may have potential clinical applications in the treatment of human spinal injuries. RGMs may also be potential targets in another axon regeneration system. Auditory neurons and their synapses with hair cells are sensitive to damage, for example from excess noise. The auditory nerve is an exception to the rule that the PNS has a greater capacity for regeneration, and auditory neurons do not exhibit this ability[131, 132, 133, 134, 135]. Recently, however, treatment with an anti-RGMA antibody was shown to increase both growth of auditory fibres and improve neuron-hair cell synaptogenesis, suggesting that RGMA antibodies could be included in a strategy to treat deafness[70].

1.4.4 Iron homeostasis disorders

There are two types of disorders in iron homeostasis, those resulting in iron overload and those causing iron deficiency. Iron overload diseases due to a genetic mutation are referred to as hereditary hemochromatosis (HH); these are disorders where iron is accumulated in vital parenchymal tissues such as the liver, heart

and pancreas and this can lead to tissue damage and fibrosis[93]. Mutations in RGMC are linked to type II HH, also called JHH, which is an autosomal-recessive early-onset form of HH that can result in severe cardiac and endocrine dysfunction as well as diabetes[136, 137, 54]. As described before, these JHH-linked RGMC mutations disrupt BMP6-mediated up-regulation of hepcidin resulting in iron overload[107, 102]. Iron deficiency disorders or anaemia are most commonly acquired by anaemia of inflammation, also known as anaemia of chronic disease. Inflammatory responses can result in low circulating iron levels, thought to be an effort to limit iron to prevent bacterial growth after infection[138]. Proteins of the inflammatory response including interleukin-6 activate hepcidin expression to prevent iron export from macrophage stores resulting in low iron availability, which can cause limited erythropoiesis[105]. Anaemia can also be caused by ectopic expression of hepcidin by tumours[100, 101] or dysregulated cytokine production in ageing[139, 140]. Interestingly, both RGMC-specific antibodies and soluble Fc-tagged RGMC have been shown to increase serum iron levels *in vivo*[103, 114]. Fc-RGMC was administered to mice over 3 weeks and resulted in decreased BMP signalling, lower hepcidin expression and corresponding increases in serum iron levels[103]. Two different antibodies targeted against RGMC were administered to rats and cynomolgus monkeys and hepcidin levels were successfully downregulated for several weeks due to the long half-life of the antibodies, with a concomitant increase in serum iron and no adverse side-effects[114]. Both groups highlight the exciting clinical potential for Fc-RGMC and anti-RGMC antibodies in the treatment of anaemia associated with hepcidin excess.

1.5 RGM signalling through NEO1

A chick RGM-alkaline phosphatase (cRGM-AP) fusion protein was used in an expression cloning strategy to search for the RGM receptor. Out of 280,000 independent clones from an adult mouse brain cDNA library expressed in COS7 cells, only one showed binding to cRGM and it was identified as mouse NEO1[141]. NEO1 was originally isolated in an antibody screen of a random cDNA library from chick cerebellum mRNA in a search for neural cell surface proteins[142]. It was immediately identified as a homologue of DCC (Deleted in colorectal cancer). DCC was originally characterised as a putative tumour suppressor gene as it was found that the deletion of one DCC allele is linked to over 70 % of human colorectal carcinomas (CRCs), and its expression is reduced in 88 % of CRC cell lines[117]. However, it was subsequently shown to be involved in axon guidance as a receptor for the chemo-attractant Netrin-1[143]. In vertebrates Netrin-1 binding to DCC triggered growth cone expansion and neurite outgrowth along a netrin-1 gradient. This was concurrently also shown in *C. elegans*, where the DCC homologue UNC-40 directed axon growth towards UNC-6 (a Netrin-1 homologue) sources, unless the UNC-5 receptor was also present when axons would be repelled[144]. As expected from its homology to DCC, NEO1 is also a receptor for Netrin-1 and appears to mediate chemo-attraction[143].

NEO1 and DCC are both members of the neuronal cell adhesion molecular family of the immunoglobulin (Ig) superfamily. This protein superfamily spans an array of functions in the immune system, the nervous system and the musculature mainly as cell adhesion and recognition molecules. NEO1 displays structural features that are conserved across the family; NEO1 is a single-span transmembrane protein with a large extracellular domain composed of four V-like Ig like domains, followed by six FN type-III like domains, with eight N-linked glycosylation sites (Fig. 1.7a). The extracellular portions of DCC and Neogenin have an

identical domain structure and 50 % amino acid identity and show particularly high similarity within the forth FN (FN4) domain (Fig. 1.7b). This domain was therefore postulated to act as the site of Netrin-1 binding as it is a common ligand of both DCC and NEO1[145]. This was recently confirmed by high-resolution crystal structures of Netrin-1 in complex with DCC and NEO1, which showed that Netrin-1 binds to the FN4-6 portions of these receptors, although the exact mode of ligand binding is still controversial[146, 147]. The cytoplasmic domains of these receptors are less conserved with only 37 % amino acid sequence identity, although three subdomains of high homology termed P1, P2 and P3 have been identified (Fig. 1.7a,b). The intracellular portion of NEO1 contains 14 potential phosphorylation sites but little is known about intracellular binding partners and hence the downstream signalling that is activated by NEO1.

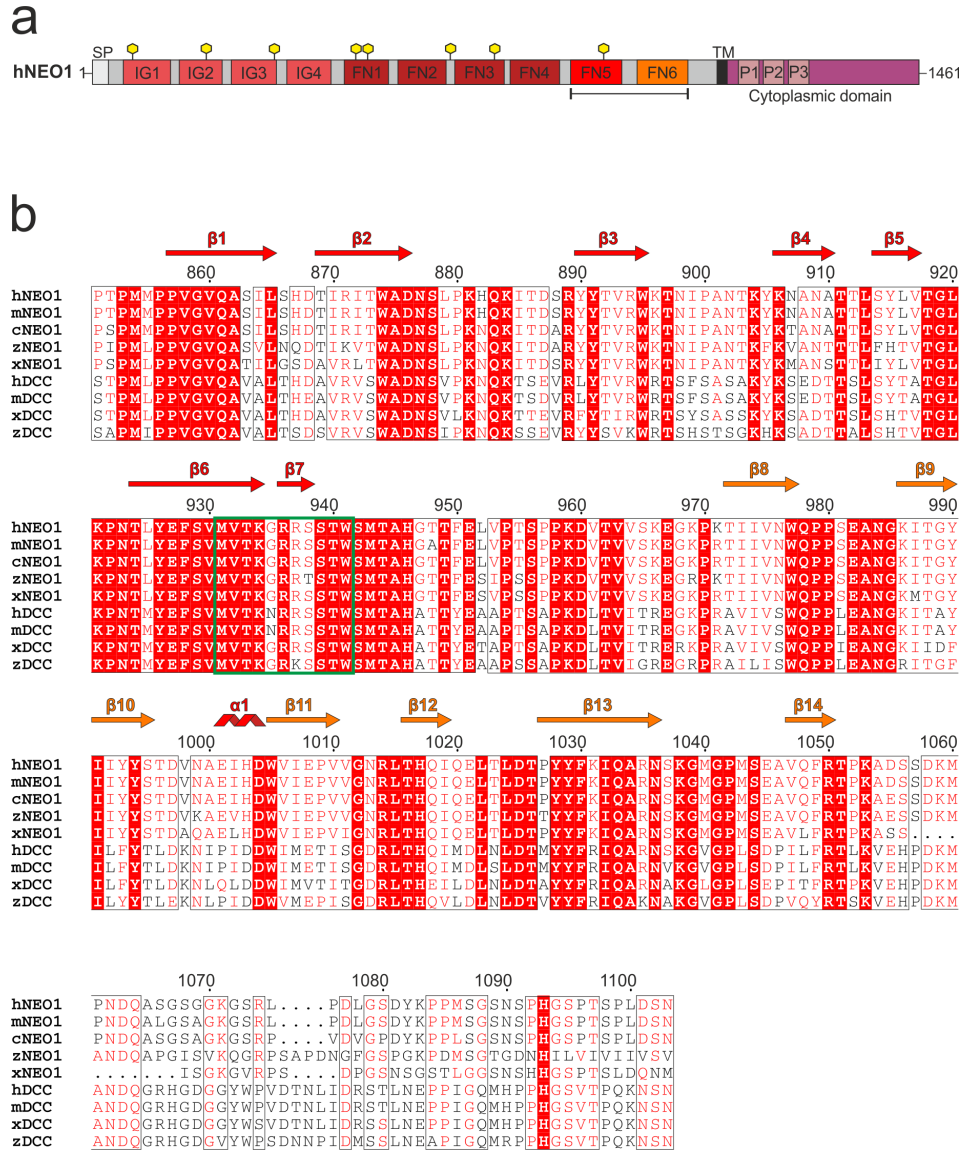
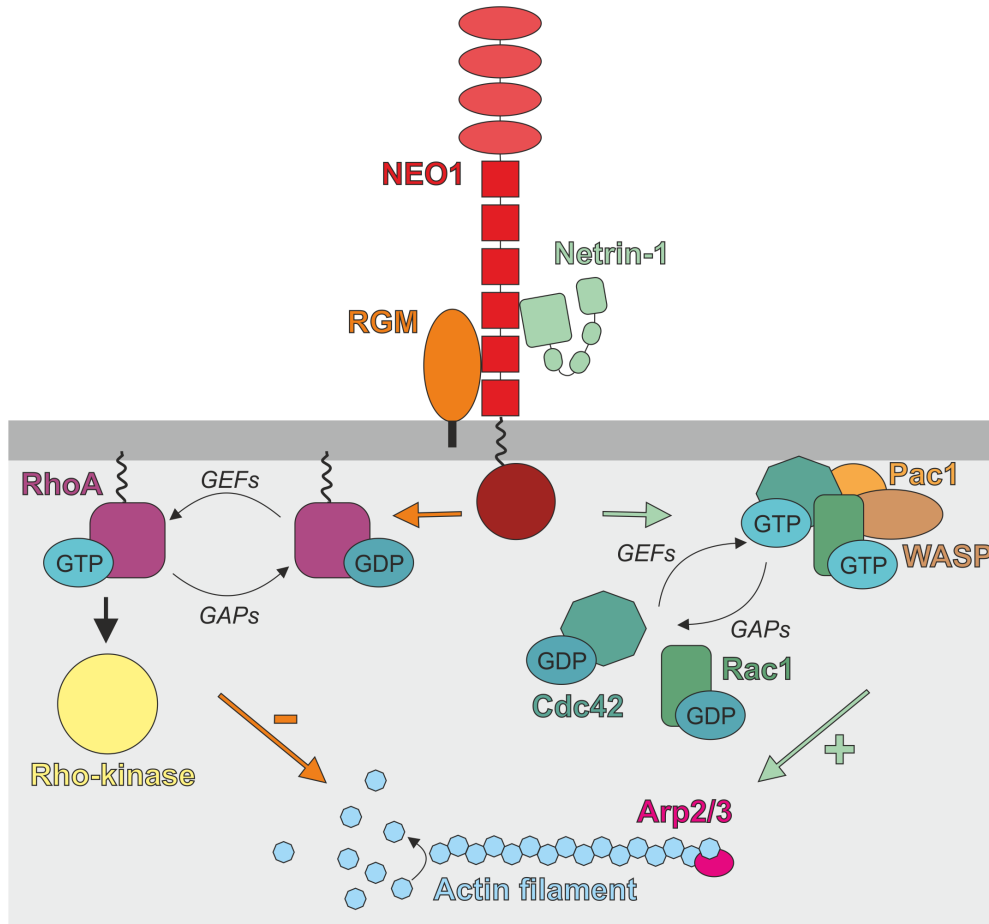


Figure 1.7: Domain organisation of NEO1 and sequence alignments of vertebrate NEO1 and DCC. From [240], Science, reprinted with permission from AAAS. (a) Schematic showing the domain organisation of human (h) NEO1. SP = signal peptide, Ig = immunoglobulin, FN = fibronectin, TM = transmembrane helix. The extracellular domain comprises four Ig domains (pale red) and six fibronectin domains (dark red, red and orange). Predicted N-linked glycans = yellow hexagons. The cytoplasmic domain (purple) comprises P1, P2 and P3 domains (pink). Sequences of NEO1 FN5-6 (black line) used in (b). (b) An alignment of vertebrate (h = human, m = mouse, c = chicken, z = zebrafish, x = *Xenopus*) NEO1 and DCC FN5-6 domain sequences. Sequences aligned using MULTALIN (bioinfo.genotoul.fr/multalin/multalin.html) and formatted with ESPRIPT (espript.ibcp.fr/ESPrpt/ESPrpt/). Numbering corresponds to hNEO1 (including secretion signal). Secondary structure assignments are displayed above the alignment, red = FN5, orange = FN6 domains[148]. The region of DCC interacting with Netrin-1[145] is boxed in green.

As described above, axon guidance cues manipulate the structure of the cell cytoskeleton by signalling through small GTPases of the Rho family including Rac1, RhoA and Cdc42 (Fig. 1.8). Netrin-1 binding to DCC has been shown to result in homodimerisation of the receptor via the cytoplasmic domain motif P3[149]. Different studies have implicated activation of different downstream signalling molecules by binding of Netrin-1 to DCC. One study showed that an intracellular signalling complex consisting of Cdc42, Rac1, Pak1 and N-WASP is recruited to DCC and results in growth cone expansion[150]. Others have demonstrated activation of either the guanine nucleotide exchange factor DOCK180, which then activates Rac1[151]; or of kinases, including focal adhesion kinase (FAK) and Fyn[152]. It is not known whether Netrin-1 binding to NEO1 results in activation of similar pathways, but the high homology of the receptor intracellular P3 domains would suggest that this is the case (Fig. 1.8). The downstream signalling triggered by binding of RGM to Neogenin is also unclear. Activation of RhoA, ROCK and protein kinase C were shown to be physiologically important for RGM signalling through NEO1[153, 67] (Fig. 1.8). Also, inactivation of FAK (also implicated in DCC-Netrin signalling) leading to activation of the GTPase activating protein p120GAP, which inactivates Ras was shown to mediate RGM-induced growth cone collapse[154].

The mechanism by which RGM binding to NEO1 triggers activation of these pathways to trigger growth cone collapse and inhibit neurite outgrowth has not yet been established. Similarly to the netrin-DCC and netrin-NEO1 interactions, the binding site for RGMs was localised to the extracellular FN domains[141]. It has since been shown that RGM binds to NEO1 at the FN5-6 domains and by analysing surface residues of these domains a potential binding site for RGM was identified, primarily on the FN6 domain, although this has yet to be confirmed[148]. Consistent with RGMs many and varied roles in development, axon guidance, immune cell regulation and iron metabolism,

Neogenin is expressed within the CNS, skeletal muscle and major organs including the heart, liver and kidneys[155, 156, 157, 158, 142]. Manipulation of NEO1 levels often mimics phenotypes associated with RGM dysfunction, for example NEO1^{-/-} knockout mice show an attenuated inflammatory response[159] and iron overload in the liver[160]. Furthermore, studies in zebrafish, using morpholinos to knockdown levels of NEO1, showed that the receptor is essential for neural tube formation[161], a finding similar to that showing that RGMA is required for neural tube closure in mice[51]. As outlined above, RGM signalling through Neogenin fulfils many roles, and has been implicated in a number of important diseases such as cancer and MS; it is therefore critical that further work focuses on enhancing our understanding of the mechanism of RGM-NEO1 signal transduction.



Actin cytoskeleton rearrangements
resulting in growth cone collapse or expansion

Figure 1.8: Two proposed pathways of RGM and Netrin-1 signalling through Neogenin. RGM and Netrin-1 bind to NEO1 at the juxtamembrane region to trigger downstream signalling. One protein thought to be activated by RGM binding to NEO1 is the RhoA GTPase, which then activates Rho kinase resulting in actin cytoskeleton rearrangements that trigger growth cone collapse. One signalling response of Netrin-1 binding to NEO1 is the recruitment of an activated complex of the small GTPases Cdc42 and Rac1 with Pac1 and WASP, which again results in actin cytoskeleton rearrangements this time enhancing growth cone expansion. GDP = guanosine diphosphate, GTP = guanosine triphosphate, GAPs = GTPase activating proteins, GEFs = guanine nucleotide exchange factors, Arp2/3 = nucleation site for actin filaments.

1.6 RGM signalling through BMPs

In addition to NEO1, the other arm of RGM signalling occurs through the BMP pathway. BMPs comprise the largest subgroup of the Transforming Growth Factor β (TGF β) superfamily with at least 16 members identified in humans. Although originally named for their ability to induce differentiation of muscle cells into bone they have since been linked to a plethora of functions from embryonic development to repair and regeneration in the adult[162, 163, 164, 165]. BMPs are synthesised as large precursor proteins consisting of a N-terminal secretion signal, a prodomain that is essential for correct folding and the C-terminal mature peptide (Fig. 1.9). The prodomain is removed by proteolytic cleavage at an Arg-X-X-Arg cleavage site by serine proteases[166, 167]. The mature BMP monomer contains seven cysteine residues, six of which form three intra-molecular disulphide bonds composing a well-characterised cysteine knot domain[168]. The mature ligands are constitutive dimers, with the seventh cysteine forming an intermolecular disulphide bond with another BMP monomer, to give the biologically active molecule[168]. BMP dimers can be homomeric or heteromeric and in some cases the prodomain remains associated with the mature dimer, adding several layers of complexity to the signalling system[169, 170, 171, 172].

BMP signalling requires binding of the ligand to a type I and a type II serine/threonine kinase receptor (Fig. 1.9). Four different BMP type I receptors (ALK1, ALK2, ALK3 (also called BMPR1A) and ALK6) and three BMP type II receptors (ACVR2A, ACVR2B and BMPR2) have been identified[162, 173, 115] (Fig. 1.10). In the activated receptor complex the constitutively active type II receptor phosphorylates the type I receptor at several serines and threonines within a highly conserved glycine- and serine-rich domain in the juxtamembrane region of the kinase known as the GS-box[174] (Fig. 1.9). This phosphorylation activates the type I receptor and creates a binding site for downstream

signalling proteins such as the receptor activated SMADs (SMAD1/5/8)[174]. The activated type I receptor phosphorylates SMAD1/5/8 and these SMADs then bind the Co-SMAD, SMAD4 (Fig. 1.10). These heteromeric complexes translocate to the nucleus and, in the presence of co-factors, bind directly to BMP responsive elements in the genome and affect gene transcription. In addition, SMAD-independent pathways can be activated by so-called non-canonical BMP signalling[175]; BMPs have been shown to activate components of the MAP kinase pathway including p38, ERK1/2 and JNK[176, 177], and also Akt/PKB, a crucial regulator of cell survival and proliferation[178]. One of the great outstanding questions of the BMP field is how so few ligands, utilising a limited set of receptors, can mediate such a diverse range of biological responses. Several mechanisms for increasing the diversity of the signalling cascade exist, including different routes of endocytosis, different modes of BMP receptor oligomerisation and a diverse set of effector proteins, both antagonists and cofactors.

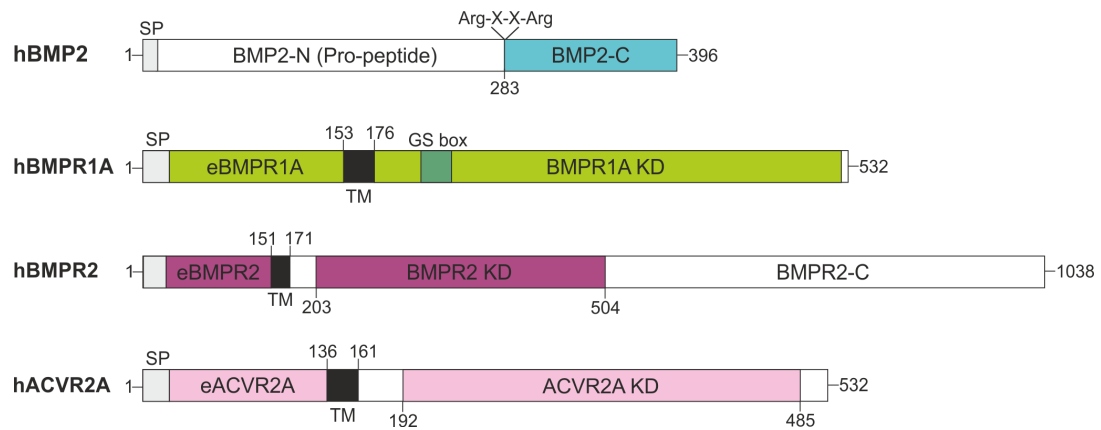


Figure 1.9: Domain organisation of the BMP ligand and its type I and type II receptors. Schematics of human (h) BMP2 and the human (h) BMP type I (BMPR1A) and type II (BMPR2 and ACVR2A) receptors. Numbering corresponds to the full-length human protein in each case. SP = signal peptide, TM = transmembrane helix, KD = kinase domain. The C-terminal portion of BMP2 making up one half of the active BMP dimer is shown in blue (BMP2-C). The ectodomains of BMPR1A (green), BMPR2 (purple) and ACVR2A (pink) are indicated with the prefix “e”. The “GS box” of BMPR1A, which is phosphorylated by the type II receptor, is shown. BMPR2 has a long C-terminal region after its kinase domain of unknown function (BMPR2-C).

The cellular localisation and the site of action of TGF β and BMP receptors are still under debate and endocytosis has been shown to be important for TGF β and BMP signalling[179]. BMP receptors (type I and II) have been shown to undergo constitutive clathrin-mediated endocytosis even in the absence of the BMP ligand, resulting in a potentiation of SMAD-dependent BMP signalling upon BMP ligand exposure[180]. Moreover, BMPRII is also internalised through caveolae and the balance between caveola- and clathrin-mediated endocytosis has been suggested to modulate the patterns of gene transcription initiated by BMP signalling[180, 181]. Inhibition of clathrin-mediated endocytosis causes a delay in SMAD1/5/8 phosphorylation and translocation of SMADs to the nucleus, and this was found to specifically reduce gene expression of a subset of BMP targets including Id1, Id3, Dlx2 and Hey1[181]. This demonstrates that endocytosis is an important regulatory mechanism in BMP signalling, adding another layer of complexity to the signalling cascade. FRET experiments have also revealed a direct interaction between both BMPRII and BMPRI and adaptor protein 2 (AP2), a key component of clathrin-coated pits, suggesting a mechanism for BMP receptor recruitment into endosomes[182]. In addition to receptor endocytosis, the BMP ligand itself has been observed to be readily internalised; confocal microscopy analysis of fluorescently labelled BMP ligand showed internalisation and accumulation towards the centre of cells in a time-dependent manner[183]. Furthermore, components of the SMAD signalling cascade are recruited to endosomal structures. The endosome-associated FYVE-domain protein endofin acts to anchor SMAD1 to the endosome for activation by BMP receptors[184]. Endofin expression was shown to enhance the phosphorylation and nuclear translocation of SMAD1 upon BMP stimulation[184]. Related to this, the endocytic adaptor protein Eps15R, a known mediator of EGF receptor internalisation, interacts with SMAD1 and depletion of Eps15R in zebrafish animal caps reduced levels of BMP signalling[185]. These lines of evidence suggest that the endosome is an important platform for BMP signalling.

Following from endocytosis, there have been many studies into the oligomeric assembly of BMP receptors in the membrane and the role that this plays in controlling downstream signalling pathway activation. The means for establishing the balance between SMAD and non-SMAD signalling is still not clear. It has been suggested that the mode of BMP receptor oligomerisation affects the downstream signalling; the BMP ligand can either bind to pre-formed type I-type II receptor complexes (PFC) or first to its high affinity type I receptor before recruiting a type II receptor in a BMP-induced signalling complex (BISC)[186]. PFCs were shown to initiate SMAD signalling, whereas BISCs activate the MAPK p38 pathway[187]. Further complexity of BMP regulation is given by multiple effector proteins that act to regulate and fine tune spatio-temporal levels of BMP signalling at the membrane[188]. These include soluble secreted antagonists (such as Noggin, Chordin and the DAN or Cerberus protein family), transmembrane proteins (e.g. BAMBI and Endoglin) and the membrane-attached RGM family (Fig. 1.10).

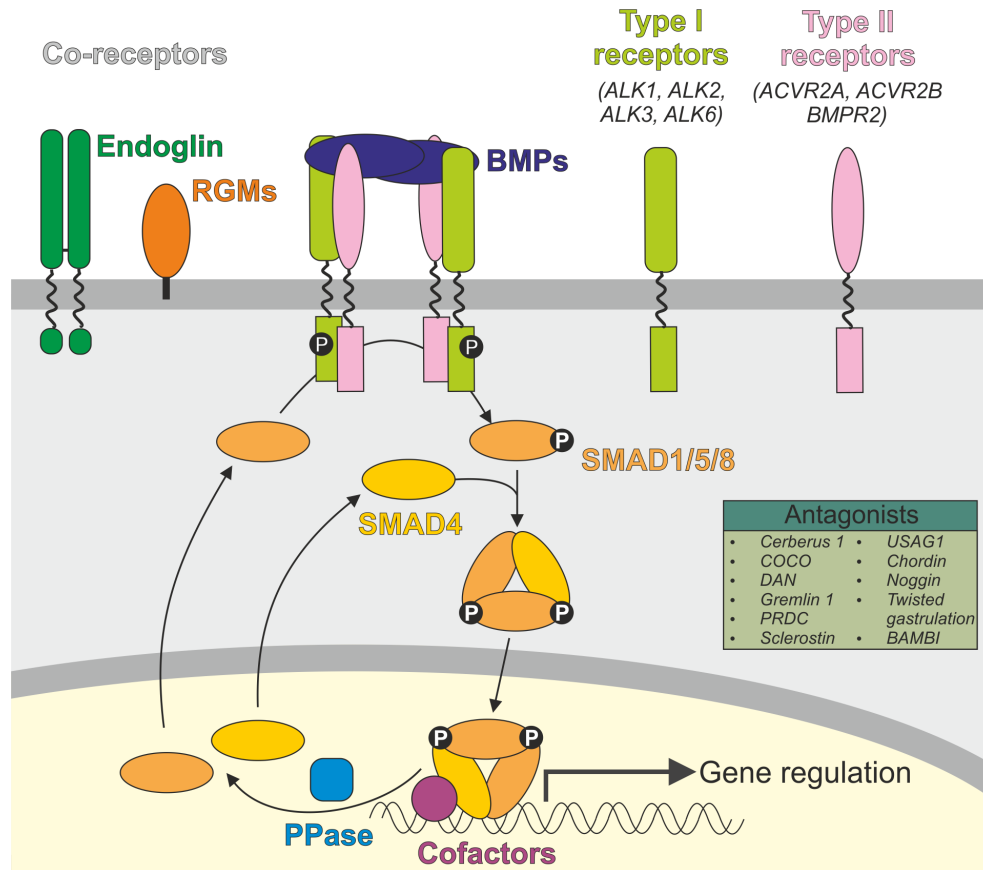


Figure 1.10: An overview of BMP signalling. Dimeric BMP ligands (dark blue) bind to type I (green) and type II (pink) receptors to form the active signalling complex. Possible receptor types are shown in brackets. The constitutively active type II receptor phosphorylates and activates the type I receptor which then phosphorylates downstream SMAD1/5/8 (pale orange). Phosphorylated SMADs form a complex with the co-SMAD, SMAD4 (yellow). This complex translocates to the nucleus and, in the presence of cofactors (purple), binds to DNA to regulate gene expression. SMAD proteins are subsequently recycled by phosphatases (PPase, blue). This signalling is enhanced by co-receptors including Endoglin (dark green) and RGMs (orange). Many different antagonists, listed in the green box, inhibit signalling. A white P in a black circle indicates phosphorylation. Figure adapted from[115].

As discussed above, all RGM family members have been shown to be important activators of BMP signalling[102, 83, 89]. This is particularly important in the regulation of iron metabolism which involves the activation of BMP6 signalling by RGM; and dysfunction of this pathway, for example through mutations in RGM, results in the blood iron overload disease JHH. While the physiological role of RGM as a co-receptor for BMP signalling has been well characterised, at least for RGM, the precise molecular mechanism by which RGM activates

this pathway is not known. All RGMs have been shown to bind directly to BMP ligands and it is known that membrane attachment of RGMs is essential. It was also shown, using co-immunoprecipitation and cross-linking studies, that RGMs can bind to all type I BMP receptors and the type II receptors ACVR2A and ACVR2B[83, 89]. Furthermore, it was shown that RGMs may affect the utilisation of the different receptor types by BMP ligands and that this might be the mechanism by which the enhanced BMP signal is generated[189]. Specifically, in KGN cells BMP ligands 2 and 4 were shown to signal exclusively via BMP type II receptor BMPR2, however, when cells were transfected with RGMA both BMPR2 and ACVR2A were utilised. An RGM-ligand-receptor complex was proposed in which RGMs enhance the affinity of the BMP ligand for a subset of BMP receptors, thereby increasing the level of BMP signalling; however, this complex is yet to be observed. Understanding the precise molecular mechanism of RGM activated BMP signalling is vital if potential therapies targeting RGMs, for example in the treatment of JHH, are to be developed.

1.7 The interplay between the NEO1 and BMP signalling pathways

RGMs signal by binding to their receptor NEO1 and also as co-receptors for BMP signalling (Fig. 1.11). However, given the overlapping expression of these proteins, particularly during embryonic development it seems unlikely that these two pathways function entirely independently of each other. Several lines of evidence have emerged in recent years demonstrating that NEO1 affects BMP signalling or *vice versa* in the presence of RGMs. The earliest indication that these two signalling pathways might be interlinked came from a study into the role of RGMC in iron homeostasis[190]. Transfection of human embryonic kid-

ney (HEK) 293 cells with RGMC increased the intracellular ferritin levels, an indicator of increased cellular iron, in comparison to control transfected cells. This effect was enhanced when the cells were co-transfected with RGMC and NEO1 and, interestingly, a JHH-linked RGMC mutation that disrupts binding of RGMC to NEO1 did not exhibit this effect[190]. Since RGMC-activated BMP6 signalling is the mechanism for iron homeostasis, this demonstrated that NEO1 might also play a role in this process. This initial finding was further developed in another study, which showed that the RGMC-NEO1 interaction is important for hepcidin expression in liver cells[191]. RGMC and NEO1 are co-expressed *in vivo* in rat hepatocytes. Knockdown of NEO1 using siRNA in RGMC-transfected HepG2 cells (human liver hepatoma cells) resulted in a 16-fold decrease in both hepcidin mRNA expression and SMAD phosphorylation, demonstrating that in the presence of RGMC, NEO1 enhances BMP signalling. Interestingly, this effect was not observed for the JHH-linked RGMC mutation G99V, which can bind to NEO1 but cannot bind to BMPs, suggesting that this activation of BMP signalling by NEO1 is mediated through the RGMC-BMP interaction. Similarly, disruption of the NEO1-RGMC interaction, using a soluble NEO1 FN5-6 fragment, disrupted RGMC activation of BMP4-induced hepcidin mRNA expression; again suggesting that NEO1 enhances BMP signalling through RGMC. Injection of this NEO1 FN5-6 fragment into mice suppressed hepcidin mRNA expression by approximately 40 %, indicating that the RGMC-NEO1 interaction is also required for the BMP-mediated induction of hepcidin *in vivo*. These results are however, in disagreement with another study, which showed in a HepB3 cell line that co-transfection of NEO1 and RGMC did not enhance hepcidin mRNA expression and down-regulation of NEO1 using siRNAs also had no effect of RGMC-mediated BMP signalling and the resultant hepcidin expression[192]. The authors state that the basis for the difference in the two results is not clear[191], although it perhaps indicates that different cell lines, with different microenvironments can exhibit different responses to RGMs.

The finding, however, that the livers of NEO1^{-/-} knockout mice exhibit severe iron overload, decreased hepcidin expression and reduced BMP signalling[160], ultimately demonstrates that NEO1 plays an important role in BMP-regulated iron homeostasis *in vivo*, most likely through RGMC.

Other reports have suggested that NEO1 regulates the shedding of RGMC from the membrane and in this way regulates the levels of BMP signalling. However, there are conflicting data on this issue with some findings indicating that NEO1 inhibits RGMC shedding[160] and others suggesting it enhances this process[193]. Interestingly, it was shown that RGMC shedding occurs from both C2C12 (mouse myoblast) and HepG2 cells suggesting that both skeletal muscle and liver could be the source of serum RGMC[193]. Knockdown of NEO1 in C2C12 cells suppressed RGMC shedding whereas overexpression of NEO1 in HEK293T cells enhanced the shedding, suggesting that NEO1 mediated RGMC release[193]. In contrast, others showed that NEO1 enhances trafficking of RGMC to the membrane and perhaps increased RGMC stability, knockdown of NEO1 resulted in accumulation of RGMC in intracellular compartments[160]. Similarly, a knockdown of NEO1 in muscle cells resulted in an increase in serum RGMC levels compared to wild type cells suggesting that NEO1 suppresses secretion of RGMC[160]. Overall, then the role of NEO1 in RGMC secretion is not clear and may differ between muscle and liver; however, regulation of RGMC shedding may be a mechanism by which NEO1 affects the levels of RGMC-activated BMP signalling in the regulation of iron homeostasis as membrane-attachment is essential for this activation and soluble RGMC has been shown to act as an inhibitor of BMP signalling.

Outside of iron regulation, evidence has been gathered for a link between NEO1 and BMP signalling in neuronal cell survival and pathfinding and in endochondral bone formation. Endochondral ossification is an essential cellular process required for the development of the skeletal system, which involves the differentiation of mesenchymal cells into chondrocytes, which eventually ossify to form

bone. BMPs are considered the master regulators of this process, and their mutation disrupts chondrogenesis[194, 195, 196]. NEO1 is also expressed in developing chondrocytes, and NEO1 mutant mice exhibit reduced skeleton size and defects in chondrogenesis and endochondral bone ossification[197]. It was found that this was linked to defective BMP signalling in NEO1 mutant mice, indicated by reduced levels of BMP target genes, including Runx2 and Id1, and a down-regulation of SMAD phosphorylation. This phenotype could be rescued by treatment with exogenous BMP2 suggesting that NEO1 may play a role in enhancing BMP signalling in chondrocytes. NEO1 appears to enhance the association of BMP receptors with membrane microdomains where BMP activates downstream SMAD signalling, a process that appears to be enhanced by RGMs[197]. Interestingly, it was shown that Neogenin is also targeted to lipid rafts in the growth cones of migrating axons, a process that appears to be necessary for RGM-induced growth cone collapse[198]. Targeting of NEO1 to these membrane microdomains seems to be dependent on BMP signalling and is abolished in the presence of the soluble BMP antagonist noggin[198]. Although it remains unclear exactly how RGM, NEO1 and BMP ligands and receptors are targeted to membrane microdomains, it appears that this process may be important for both effective axon guidance and BMP signalling, further supporting the hypothesis that these fundamental signalling pathways are inter-reliant. Furthermore, evidence that the relationship between these pathways is conserved through evolution was provided by work carried out in the model organism *C. elegans*[199]. The NEO1 homolog UNC40 was shown to enhance a BMP-like signalling pathway in a process that was dependent on the association of UNC40 with DRAG-1, an RGM homolog. Mutations in UNC40 and DRAG1 resulted in developmental patterning phenotypes typically associated with mutants in the Sma/Mab (BMP-like) pathway. Interestingly these phenotypes could be partially rescued by expression of mouse Neogenin and human RGMC, demonstrating the level of evolutionary conservation of both proteins and their involvement in BMP

signalling[199]. Taken together, there is a wealth of functional evidence suggesting that NEO1 and BMP signalling are linked and that this is important for the correct function of both pathways. This interaction appears to be mediated by RGMs, which bind directly to both the NEO1 receptor and BMP ligands. However, the architecture of this super-complex is not known and this physical link has yet to be demonstrated (Fig. 1.11). It is important to understand the interplay between these two fundamental pathways as a full understanding of their inter-dependence may open up new therapeutic targets in the treatment of diseases that feature dysfunction of either pathway as their basis.

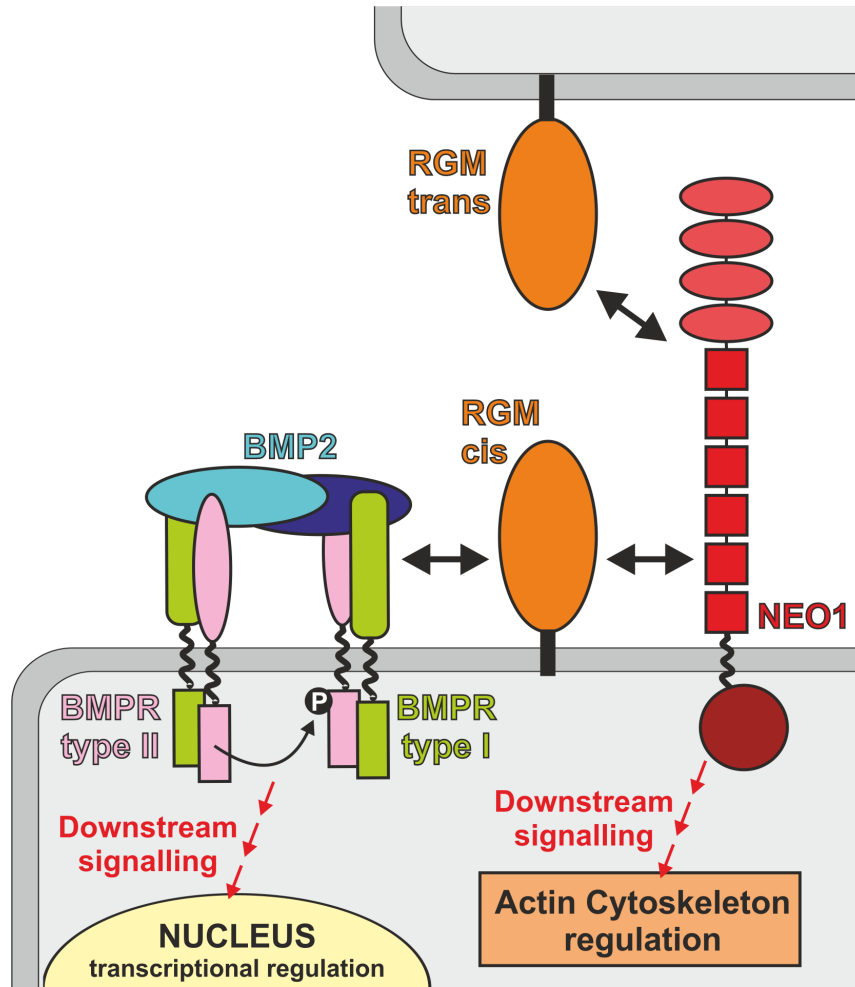


Figure 1.11: RGMs signal through two fundamental signalling pathways. RGMs (orange) can signal in *cis* (same-cell) or *trans* (cell-to-cell), which can include shedding of the RGM to produce a soluble form with long-range actions. *Trans*-signalling occurs through NEO1 (red shades) to trigger actin cytoskeleton rearrangements and subsequent growth cone collapse. Signalling in *cis* can occur via activation of the BMP signalling pathway, to effect changes in gene expression, and also via NEO1. The form of RGM interaction with the BMP pathways is not clear it has been shown to bind to both BMP ligand (blue) and receptors (green and pink). The BMP and NEO1 signalling pathways are linked, supposedly by RGM; however, a physical link has yet to be demonstrated.

1.8 Thesis outline

RGM family members fulfil a plethora of roles within the developing embryo and the adult, which cannot be explained by a single mode of signalling. RGMs signal through the NEO1 and BMP pathways to trigger different downstream signalling cascades, however, the precise molecular mechanisms of RGM signalling are not known. The mode of RGM binding to NEO1 or to the BMP ligand has not been determined, and the architectures of the active signalling complexes in either case have not been defined. Furthermore, there is a great deal of functional evidence supporting the interaction between the NEO1 and BMP pathways, which has been suggested to be mediated by RGM. However, evidence of RGM acting as a physical link between the NEO1 receptor and a component of the BMP pathway has not yet been demonstrated. In order to address these outlying questions X-ray crystallography was used to determine the structures of complexes between RGMs and components of these signalling pathways. In Chapter 2 of this thesis, the structural, biophysical and cellular methods used throughout this project to investigate the molecular mechanisms of RGM signalling are described. The primary method utilised in this work is protein X-ray crystallography, and the complex structures solved guided all subsequent experiments. Biophysical techniques were also employed, ranging from small angle X-ray scattering to analytical ultracentrifugation to surface plasmon resonance, which were used largely to investigate the oligomeric forms of protein complexes and to validate complex interfaces observed in the structures. This then led to the use of cellular techniques including luciferase reporter assays and super-resolution fluorescence microscopy, which were employed to begin to test the functional relevance of the protein complex structures.

In Chapter 3 a structural analysis of an RGMB-NEO1 structure that was previously solved in our laboratory is discussed. This analysis guided mutagenesis

studies that were subsequently used in analytical ultracentrifugation experiments to demonstrate that a 2:2 stoichiometry form of the RGMB-NEO1 complex exists in solution. Mutant RGMB proteins were expressed and purified for use in a neurite outgrowth collapse assay, which was carried out by our collaborators in the Pasterkamp laboratory, and demonstrated that the 2:2 complex form is physiologically important for RGM signalling through NEO1. Full-length NEO1 receptors were also shown to cluster in the membrane, which led to the formation of a hypothesis for RGM signalling through NEO1: RGM binds to pre-clustered NEO1 at the FN5-6 domains of the receptor and stabilises the orientation of the domains thus triggering downstream signalling.

In chapter 4, the crystallisation and structure solution of complexes of the N-terminal domain of all RGM family members in complex with the BMP2 ligand is detailed. An analysis of this interface, followed by mutagenesis and surface plasmon resonance binding studies, verified the binding mode observed in the structure. Furthermore, this analysis informed a potential molecular mechanism for JHH-linked mutations in RGMC that occur at the binding interface with BMP, as these mutations disrupted the binding of the ligand to its co-receptor RGMC. A comparison between the structures of the RGM-BMP and BMP-BMP type I receptor ectodomain complexes showed that RGM and BMPR1A share a binding site on the BMP ligand. The two proteins compete for binding to the BMP ligand and behave differently in cellular luciferase reporter assays measuring BMP signalling. Furthermore, multi-angle light scattering experiments showed that the RGM-BMP interaction is pH-dependent, unlike the BMP-BMPR1A complex, and from this information an endocytosis-linked mechanism for RGM-activated BMP signalling was proposed, which is linked to the BMP receptor subcellular localisation. Several fluorescence microscopy experiments were performed in an attempt to gather evidence in support of this model; but these were unsuccessful due to technical difficulties, and this is discussed. Altogether this work provides

the first structural information for the RGM-BMP interaction and begins to shed light on the mechanism for RGM-activated BMP signalling.

In chapter 5, the crystallisation and structure solution of a ternary BMP2-RGMB-NEO1 complex is described. In this complex the interfaces identified in the RGM-NEO1 and RGM-BMP complexes are replicated, with the linker between the RGM N- and C-terminal domains connecting the two complexes. Small-angle X-ray scattering data of the ternary complex showed that the complex architecture observed in the crystal structure is preserved in solution. In the complex the C-termini of both RGMB and NEO1 are oriented in the same direction, presumably toward the plasma membrane. This complex, combined with the 2:2 stoichiometry RGMB-NEO1 complex, shows that NEO1, RGM and BMP could associate into clusters on the membrane. Total internal reflection, direct stochastic optical reconstruction microscopy experiments showed that the addition of BMP2 resulted in time-dependent clustering of fluorescently tagged NEO1, a process that was dependent on the presence of RGMB. This BMP2-induced clustering of the NEO1 receptors could be important in a multitude of signalling processes. Taken together these data show that RGM acts as a bridge between NEO1 and BMP, demonstrating, for the first time, that RGM physically links these two fundamental signalling pathways.

In chapter 6 I will discuss the implications of these findings with respect to the wider literature. The results presented in this thesis naturally lead to many future experiments that could vastly improve our understanding of these complicated signalling systems. Of particular interest is the physiological relevance of the structures presented herein, and of the clustering hypothesis resulting from the ternary complex. It will be very interesting to probe the role of this in NEO1 and BMP signalling *in vivo*. This work assimilates structural studies of RGM signalling through its receptor NEO1, and RGM acting as a co-receptor in BMP signalling. This structural data, in combination with biophysical and cellular

studies begins to shed light on the molecular mechanisms of RGM signalling through these two pathways, showing that RGM is a structural bridge connecting these two essential developmental pathways. This structural data could guide the design of novel therapeutics in the treatment of RGM-linked diseases such as multiple sclerosis and JHH, where RGMs are just beginning to be explored as viable disease treatment targets.

Chapter 2

Materials and Methods

2.1 Molecular Biology

2.1.1 cDNA and expression vectors

For expression in bacteria the human bone morphogenetic protein 2 mature peptide (termed BMP2, residues Q283-R396 of BMP2, GenBank ID AF040249) was cloned into the bacterial expression vector pET22B (Invitrogen) containing the additional start codon ATG, resulting in an N-terminal methionine residue. For expression in mammalian cells, constructs of the extracellular region of human RGMA (GenBank ID AL136826, eRGMA: P47-Y414, RGMA_{ND}: T45-S139), human RGMB (GenBank ID AK074887, eRGMB: Q50-S410, RGMB_{ND}: Q53-H136, RGMB_{ΔND}: D137-S410), human RGMC (GenBank ID AY372521, eRGMC: Q36-S399, RGMC_{ND}: Q36-P147), human BMPR1A (GenBank ID AK291764, eBMPR1A: T49-V141, BMPR1A-mVenus: T49-I532), human BMPR2 (GenBank ID Z48923, eBMPR2: L33-P132, BMPR2-mTFP: S27-L1038, BMPR2-mTFP: S27-E509) and mouse NEO1 (GenBank ID Y09535; NEO1: 37-1493, NEO1_{FN56}: 883-1083), were cloned into either the pHLsec vector, which adds a C-terminal His₆ tag, or one of a suite of pHLsec variant vectors, which include a C-terminal biotin ligase (BirA) recognition sequence, or a 3C protease linker (LEVLFQGP) followed by an mVenus or mTFP fluorescent protein[200]. The pHLsec vectors contain an optimised Kozak sequence, to facilitate the initiation of translation, a secretion signal sequence (MGILPGMAPALLSLVSLLSVLLMGCVA) and AgeI and KpnI restriction sites to allow for easy insertion of constructs into the pHLSec vector.

2.1.2 Construct design and PCR

Expression constructs were designed using the homology-based protein structure prediction software HHpred[201], the protein disorder prediction software RONN[202] and the signal peptide prediction software SignalP[203]. Protein sequence annotations provided by the UniProt webserver[204] were also used to search for conserved regions of sequence in protein orthologues and homologues; the sequence alignments were performed using Multalin[205]. PCR reactions were carried out in a 50 μ l reaction volume containing: 1 μ l each of forward and reverse primers (10 pmol/ μ l), 1 μ l template DNA (less than 100 ng), 4 μ l dNTP mixture (2.5 mM) (Takara), 5 μ l 10 X PyrobestTM Buffer II (Takara) and 0.25 μ l PyrobestTM DNA Polymerase (5 U/ μ l) (Takara) made up to 50 μ l with MilliQ[®] H₂O. PCR primers were purchased from Eurofins (Eurofins Genomics[®]). PCR cycles were carried out in a T100 thermal cycler (BIORAD) as follows:

1. 98 °C for 10 s (denaturation of DNA double strands).
2. 98 °C for 10 s (as above).
3. 55–60 °C for 30 s (annealing of primers to DNA template).
4. 72 °C for 60 s + 60 s/kb PCR product (elongation of DNA strands).
5. Repeat steps 2–4 30 times.
6. 72 °C for 600 s (ensure complete elongation).
7. 4 °C for ∞ .

PCR products were either purified using the QIAquick PCR Purification Kit (Qiagen) or loaded onto a 1 % agarose gel, excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen).

2.1.3 Restriction digests, ligations and transformations

Purified PCR products and 2 μg of the required vector were digested with the required restriction enzymes using the recommended buffer conditions for 2 hours at the appropriate temperature (New England Biolabs). Digested products were then loaded on a 1 % agarose gel in Tris-acetate-EDTA (TAE) buffer. Bands of the expected size were excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen). Ligations reactions were carried out using the Quick LigationTM Kit (New England Biolabs) in a final volume of 20 μl according to the manufacturers instructions, using a vector to insert molar ratio of 1:4. Subsequently, 1 μl of this ligation mixture was used to transform 20 μl of Library Efficiency[®] DH5 α TM competent cells (Life Technologies). The DNA-cell mixture was incubated on ice for 30 min before being heat-shocked at 42 °C for 45 s. After incubation on ice for 2 min, 100 μl Super Optimal Broth (SOC) medium was added and the mixture was incubated at 37 °C for 45 min with 200 rpm shaking. Cells were then spread onto LB-Agar plates containing 100 $\mu\text{g}/\text{ml}$ ampicillin and incubated at 37 °C overnight. Two colonies were then picked to inoculate overnight cultures (5 ml) for purification of the plasmid DNA using the QIAprep Spin Miniprep Kit (Qiagen). Clones were then sequence-verified (Source Bioscience) and stored at -20 °C.

2.1.4 Site-directed mutagenesis

Site-directed mutagenesis of constructs, for example to introduce point mutations, was carried out using a two-step overlap extension PCR method. This method required four primers: one each to define the 5' and 3' ends of the construct, and a pair of complementary primers that spanned the target site (approximately 21 bp length) and harboured the required mutation at their centres. In the first step, two PCR reactions were carried out resulting in two PCR prod-

ucts comprising the sequence from the mutation site to both the 3'-terminus and the 5'-terminus. In the second PCR step, the gel-purified PCR products from the first step were used as the template for this reaction (1 μ l each) along with the primers defining the 3' and 5' termini. The product of this second reaction spanned the entire length of the expression construct and contained the required mutation. This DNA fragment was then gel-purified and cloned into the required vector.

2.2 Western Blotting

Proteins were separated according to their size by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (120 V, 45 min) using 4-12 % Bis-Tris PAGE gels (Lifetechnologies). Proteins were then blotted onto a nitrocellulose membrane (30 V, 90 min); the membrane was blocked (5 % milk in 0.01 M phosphate buffer, 0.0027 M potassium chloride and 0.137 M sodium chloride, pH 7.4 (1 X phosphate buffered saline (PBS), from tablets, Sigma), 30 min), and probed with a primary antibody for 1 hour at room temperature; primary antibodies used were mouse anti-His₆ (1:1000 dilution, Qiagen), mouse anti-1D4 (1:2000, University British Colombia) and rabbit anti-Phospho-Smad1/5 (Ser463/465) (1:1000, Cell Signaling Technology[®]). Following a brief wash step in 1 X PBS the membrane was probed with a horseradish peroxidase (HRP)-conjugated secondary antibody for 1 hour at room temperature: either goat anti-mouse (1:2000 dilution, Sigma) or goat anti-rabbit (1:2000, Cell Signaling Technology[®]). The membrane was washed 3 times for 15 min in 1 X PBS, 0.1 % Tween 20 and the signal was detected using an enhanced chemiluminescence (ECL) kit (GE Healthcare) by exposure onto Hyperfilm ECL (GE Healthcare).

2.3 Protein expression and purification

2.3.1 Expression and purification of BMP2 from *E. Coli*.

The pET22B vector containing the BMP2 sequence (2 μ l) was transformed into 50 μ l RosettaTM 2(DE3) SinglesTM competent cells (Novagen) for expression in *E. Coli*. Transformed cells were spread onto LB-Agar plates containing 100 μ g/ml ampicillin, 34 μ g/ml chloramphenicol and 1 % glucose. Single colonies were picked to set up two overnight starter cultures (100 ml) in LB broth containing 100 μ g/ml ampicillin and 34 μ g/ml chloramphenicol. These were then used to inoculate 6 L of LB broth containing 100 μ g/ml ampicillin (13 ml overnight culture/L). Cultures were grown at 37 °C in a shaking incubator (200 rpm) to an A₅₉₅ of 0.8, BMP2 expression was induced by addition of 1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) and the cultures were incubated at 37 °C for 6 hours. Cells were harvested by centrifugation (7000 xg, 4 °C, 10 min) and pellets were frozen overnight at -80 °C. BMP2 was expressed as inclusion bodies and purified as follows, using a protocol adapted from[206]. Bacterial pellet from 1 L of culture was resuspended in 40 ml 20 mM EDTA, 2 % (v/v) Triton X-100, 500 mM NaCl, pH 7, supplemented with 1 ml of protease inhibitor cocktail (P8465, Sigma Aldrich) and 500 μ l DNase (Sigma Aldrich). Cells were lysed by sonication (10 min, amplitude 50 %) (Sonics VCX500) and the lysate was cleared by centrifugation (31,000 xg, 8 °C, 45 min). After cell lysis, the inclusion body pellets were washed four times with 20 mM EDTA, 2 % (v/v) Triton X-100, 500 mM NaCl, pH 7 by repeated rounds of sonication (10 min, 50 %) and centrifugation (10,000 xg, 8 °C, 10 min). The resultant inclusion body pellet was solubilised in 6 M guanidinium hydrochloride (GdmHCl), 0.1 M Tris-HCl, pH 8.5, 1 mM EDTA and 100 mM DTT overnight at 4 °C. The pH was dropped to 3-4 to inhibit disulphide bond formation by drop-wise addition of HCl. Residual insoluble material was removed by centrifugation (31,000 g, 8 °C, 10 min). DTT

was removed by dialysis against 10- to 20-fold volume of 6 M GdmHCl, pH 3-4. The dialysis step was repeated at least four times. Refolding was then carried out by incubating the reduced and solubilised inclusion body preparation for 24 hours at 4 °C at a concentration of 200 μ g/ml in 100 mM Tris-HCl, 5 mM EDTA, 1 M L-arginine, pH 8.3, 100 μ M oxidised glutathione (GSSG), 100 μ M reduced glutathione (GSH). The protein concentration was increased by concentrating the sample to 1 mg/ml using Amicon pressure filtration (Millipore). The reaction was split into two halves, one of which was oxidised with 25 mM GSSG for 3 h at 4 °C. After this incubation, excess GSSG was removed by dialysis and the two reaction halves were combined. The sample was then dialysed against 100 mM Tris-HCl, 5 mM EDTA, 6 M Urea, pH 6.0 and applied to a heparin column (5 ml HiTrap Heparin HP, GE Healthcare). Bound protein was eluted over a NaCl concentration gradient (from 10 mM to 1 M NaCl) over 35 column volumes. Fractions containing the BMP2 dimer were further purified by size exclusion chromatography (SEC) in 4 M Urea, 100 mM Tris-HCl, pH 6.0, 5 mM EDTA. SEC and heparin column fractions were analysed by non-reducing SDS-PAGE and those with a purity exceeding 95 % of the dimeric species were pooled.

2.3.2 Small-scale expression trials in mammalian cells

Constructs of RGM, NEO1 and BMP receptor proteins were cloned into the pHLSec vector[200] for large-scale expression. However, small-scale expression tests were first conducted to ensure the constructs were secreted. Human embryonic kidney (HEK) 293T cells were maintained in complete Dulbecco's Modified Eagles Medium (DMEM) supplemented with non-essential amino acids (Gibco), L-glutamine and 10 % Fetal Bovine Serum (FBS, Sigma), at 37 °C, 5 % CO₂. For small-scale trials cells seeded at a density of 1 x 10⁵ cells/ml in 6-well plates and grown to approximately 90 % confluency. The media was then exchanged

for DMEM supplemented with 2 % FBS (1.5 ml). Transfection cocktail was prepared by mixing 4 μ g DNA in 250 μ l serum-free DMEM with 8 μ l Lipofectamine 2000 transfection reagent (Lifetechnologies) in 250 μ l serum-free DMEM. The Lipofectamine-DNA mixture was incubated at room temperature for 20 min and then added to the cells. The cells and media were harvested three days after transfection and analysed by Western blotting using a mouse anti-His₆ antibody probe to detect secreted proteins.

2.3.3 Expression and purification of secreted proteins from mammalian cells

Constructs of RGM, NEO1 and BMP receptor proteins were cloned into the pHLSec vector[200] for large-scale expression by transient transfection of mammalian HEK293T cells using a previously described semi-automated procedure[207]. The large amounts of DNA required for large-scale transient transfection of HEK293T cells were prepared using the Endotoxin-Free Plasmid High Speed Giga Kit (Qiagen). For this, 2 L of bacterial culture (*E. Coli*, DH5 α cells) transformed with the required plasmid construct were grown at 37 °C overnight and the cells were harvested by centrifugation (3,000 xg, 10 min). One Giga-Prep resulted in approximately 12-15 mgs of purified plasmid DNA. Briefly, transient transfections were performed using HEK293T cells. These were grown in complete DMEM supplemented with 10 % FBS in 2125 cm² expanded surface roller bottles (Greiner). Chloroform (80 μ l/ml) was added to the plasmid DNA (0.5 mg/roller bottle) and the mixture was centrifuged (16,000 xg, 10 min) to remove impurities. Purified plasmid DNA was added to serum-free DMEM (50 ml/roller bottle) and the transfection reagent polyethylenimine (PEI) was then added to the DNA mix (1 mg/roller bottle). This solution was incubated for 10 min at room temperature to allow for PEI-DNA complex formation. During

this incubation period the media from the roller bottles was exchanged for complete DMEM supplemented with 2 % FBS (200 ml/roller bottle). The class I α -mannosidase inhibitor kifunensine[208] was then added to the DNA-PEI mixture (0.25 mg/roller bottle). Kifunensine inhibits an early step in the mammalian glycosylation pathway, resulting in the addition of homogenous oligomannose-type N-glycans with the general structure Man5-9GlcNAc2 (Man, mannose and GlcNAc, N-acetyl glucosamine), which are sensitive to enzymes such as endoglycosidase F1 (Endo-F1)[208, 209]. Endo-F1 cleaves between the GlcNAc residues in the di-N-acetylchitobiose core of oligomannose sugars, reducing the N-glycans to single residues thereby enhancing crystallisation[209]. Finally the DNA-PEI mixture was added to each roller bottle. Cells were incubated at 37 °C in a roller rig (Wheaton) for 4-6 days before the media was harvested. Cells were removed by centrifugation (7,000 xg, 20 min) and the media was filtered through a 0.2 μ m sterile vacuum filtration device (Millipore). The filtered conditioned medium was dialysed (for 48 hours at 4 °C) against 25 mM PBS (23 mM Na₂HPO₄, 1.7 mM NaH₂PO₄, 250 mM NaCl, pH 7.5) and the proteins were purified by immobilised metal-affinity chromatography using TALON beads (Clontech). Beads were washed with 5 mM Imidazole in 25 mM PBS and eluted in 350 mM imidazole in 25 mM PBS. Proteins were then purified by SEC (Superdex 200 16/60 column, GE Healthcare) in buffer containing 10 mM HEPES, 150 mM NaCl, pH 7.5.

2.3.4 Co-immunoprecipitation of trans-membrane proteins

Transfection and co-transfection of full-length NEO1-His₆ and NEO-1D4 plasmids, in HEK293T cells grown in roller bottles, were performed as described above. About 40 hours post-transfection the media (250 ml) was removed. The cells were washed in 2 x 50 mL PBS (pH 8.0) and lysed in 5 mL Lysis Buffer (10

mM Tris-HCl (pH 8.0), 150 mM NaCl, 1.5 % DDM and protease inhibitor cocktail (Sigma)) for 2 hours at 4 °C. The lysate was cleared by centrifugation (100,000 xg, 2 hours) and diluted 1:50 in Lysis buffer containing 0.03 % DDM. The concentration of detergent was lowered to allow for NEO1 complex formation. Complexes were recovered on 1D4-antibody (University British Colombia)-conjugated sepharose beads (Amersham) and washed three times with Lysis buffer (0.03 % DDM). Bound protein complexes were eluted from the beads by incubation (12 hours, 4 °C) with 500 μ M 1D4 peptide (TETSQVAPA, GenScript). Samples were analysed by Western blotting with a mouse anti-His₆ antibody probe.

2.4 Protein crystallisation and data collection

Prior to crystallisation, complexes were concentrated (RGMA_{ND}-BMP2: 6 mg/ml, RGMB_{ND}-BMP2: 6 mg/ml, RGMC_{ND}-BMP2: 5 mg/ml and BMP2-eRGMB-NEO1_{FN56}: 5 mg/ml) and treated with Endo-F1 (75 μ g mg⁻¹ protein, 1 h, 37 °C) to cleave glycosidic bonds of N-linked sugars resulting in only one N-acetylglucosamine moiety bound to the corresponding asparagine side chain. Crystallisation trials, using 100 nl protein solution plus 100 nl reservoir solution in sitting drop vapour diffusion format were set up in 96-well Greiner plates using a Cartesian Technologies robot[210]. Crystallisation plates were maintained at 6.5 or 20.0 °C in a TAP Homebase storage vault and imaged via a Veeco visualisation system[211]. All binary RGM-BMP2 complex crystals were grown at 6.5 °C, whereas the ternary BMP2-RGMB-NEO1 complex was crystallised at 20 °C. RGMA_{ND}-BMP2 crystals were obtained out of mother liquor containing 0.1 M citric acid, pH 4.0, 20 % (v/v) 2-methyl-2,4-pentanediol (MPD), 0.2 M glycine, RGMB_{ND}-BMP2 Form-1 crystals out of 2 M ammonium sulphate, 8 % (v/v) 2,5-hexanediol, RGMB_{ND}-BMP2 Form-2 crystals out of 0.08 M citric acid, pH 4.0, 15 % (v/v) MPD, RGMC_{ND}-BMP2 crystals out of 20 % (v/v) PEG3350, 0.2 M am-

monium nitrate, eBMPr1A-BMP2 crystals out of 0.1 M Tris-HCl, pH 8.0, 65 % (v/v) MPD, 0.4 M NDSB-221 (non-detergent sulfobetaines) and BMP2-eRGMB-NEO1_{FN56} crystals out of 0.1 M Tris-HCl, pH 8.5, 1.5 M ammonium sulphate, 12 % (v/v) glycerol. Prior to flash-freezing, crystals were treated with the appropriate cryo-protectant solutions (RGMA_{ND}-BMP2 and RGMB_{ND}-BMP2 Form-2: 30 % (v/v) MPD in mother liquor; RGMB_{ND}-BMP2 Form-2 and RGM_{CND}-BMP2: 30 % (v/v) ethylene glycol in mother liquor; BMPr1A-BMP2: due to the high percentage MPD in the mother liquor, no cryo-protectant was added; BMP2-eRGMB-NEO1_{FN56}: 15 % (v/v) glycerol in mother liquor. Diffraction data were collected at 100 K at beamlines I03 (RGMB_{ND}-BMP2 Form-1, RGM_{CND}-BMP2, BMP2-BMPr1A, BMP2-eRGMB-NEO1_{FN56}), I04 (RGMA_{ND}-BMP2) and I04-1 (RGMB_{ND}-BMP2 Form-2) at the Diamond Light Source, UK. Data collection strategies were guided by EDNA[212] and MOSFLM[213] as part of the automatic software pipeline installed at Diamond. Data collection statistics for all structures solved in this thesis are given in Table 1.

2.5 Basic principles of biomolecular crystallography

2.5.1 Protein crystal symmetry

Protein molecules can pack in a regular periodic arrangement, or lattice, to form a crystal that results in X-ray diffraction. The smallest volume that can be infinitely repeated to reproduce the entire crystal is called the unit cell, which can be described by three vectors **a**, **b** and **c**; and three angles α , β and γ (Fig. 2.1a). Each unit cell in the crystal contains the same number of identically arranged protein molecules, which are related by internal symmetry operations including

rotational and screw axes. There are 230 possible crystal-packing arrangements or space groups, but protein crystals are limited, due to their chirality, to only 65 of these. The asymmetric unit of a crystal is the minimum motif, which when combined with the crystallographic symmetry operations (including unit cell parameters and crystal space group), can be used to reconstruct the protein crystal. The space groups and unit cell dimensions for all the structures solved in this thesis are given in Table 1.

A unit cell can be further divided by sets of parallel lattice planes, separated by an inter-planar distance (d_{hkl}) and these are defined by the Miller indices[214] h , k , and l (Fig. 2.1b). Higher values for the Miller indices indicate a smaller inter-planar distance (Fig. 2.1c,d). Diffraction off these lattice planes using X-rays gives information about the contents of the unit cell.

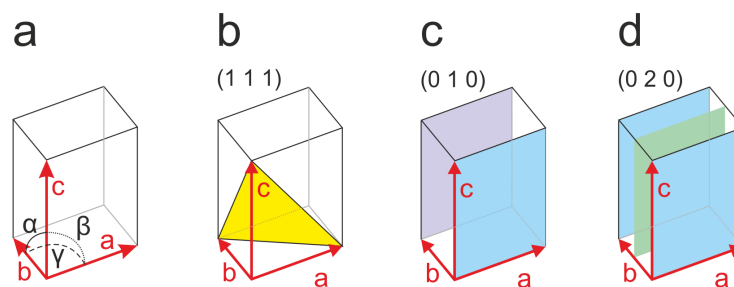


Figure 2.1: The unit cell and Miller indices. (a) A unit cell can be described by three vectors ($\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$) and three angles (α , β , γ). (b) The hkl plane with Miller indices $(1\ 1\ 1)$ is shown, the plane intersects each unit cell vector once within the unit cell. (c) The hkl plane with Miller indices $(0\ 1\ 0)$, the plane intersects b once but never intersects a or c . (d) The hkl plane with Miller indices $(0\ 2\ 0)$, the plane intersects b twice. Note that the $(0\ 1\ 0)$ plane is also a member of the set of $(0\ 2\ 0)$ planes, they belong to the same zone of planes. Figure adapted from Figure 5-48[215].

2.5.2 X-ray diffraction

X-rays are electromagnetic waves with wavelengths of approximately $1\ \text{\AA}$, they are composed of an amplitude and a phase (Fig. 2.2a). X-rays are scattered by

electrons and therefore the total scattering power of an atom is proportional to its number of electrons. In a crystal, X-rays scattered from different families of lattice planes will have different phases due to the different path lengths, d_{hkl} , between the planes, and the amplitudes of the resultant scattered waves will be affected by constructive and destructive interference (Fig. 2.2a). If the path length between the lattice planes is an integer multiple of the X-ray wavelength then constructive interference reaches a maximum, this is known as the Laue condition[216, 217]. X-rays can be described as vectors comprising an amplitude ($\mathbf{F}$) and phase angle (φ) and depicted graphically using an Argand diagram[218], which allows wave summation to be conveniently reduced to vector addition (Fig. 2.2b).

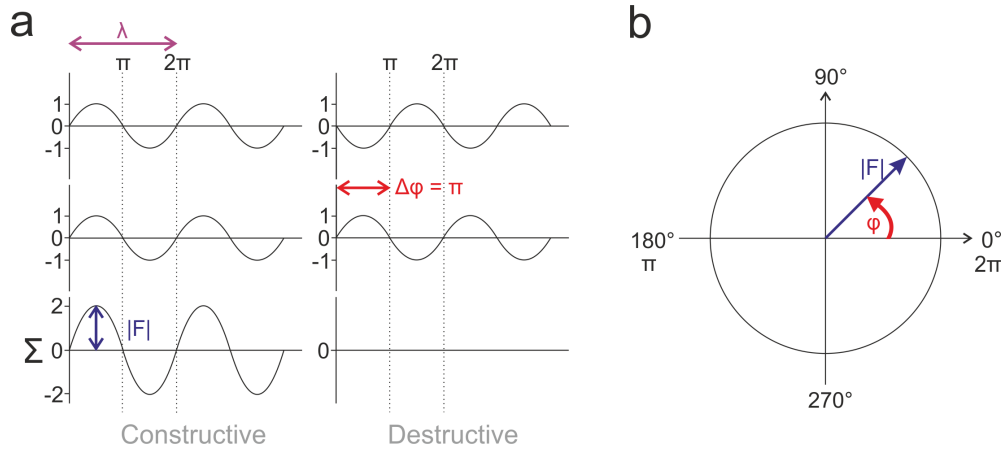


Figure 2.2: Summation of waves. (a) Waves are made up of an amplitude ($|\mathbf{F}|$, shown in blue) and a phase (φ), here the phase difference ($\Delta\varphi$) between two waves is shown in red. Constructive interference, where the waves are in phase, results in an increase in amplitude, whereas destructive interference, where waves are out of phase, results in a decrease in amplitude. (b) An Argand diagram represents waves as vectors in the complex number plane (an example wave vector is shown in blue). The magnitude of the vector represents the amplitude ($|\mathbf{F}|$) of the wave and the orientation of the vector in space, indicated by the phase angle (φ), represents the phase. Figures adapted from Figure 6-4 (a) and Figure 6-5 (b)[215].

2.5.3 Bragg's Law

Bragg's Law (Equation 2.1) describes a way to define X-ray diffraction by a crystal in terms of the wavelength of the incident light (λ), the scattering angle (θ) and the distance between lattice planes (d_{hkl}) [219] (Fig. 2.3a).

$$n\lambda = 2d_{hkl} \sin \theta \quad (2.1)$$

To satisfy the Laue condition and result in an observable diffraction maximum (recorded as a diffraction spot) d_{hkl} must equal an integer multiple of the wavelength. Importantly, this means that at a fixed wavelength the positions of the diffraction maxima are dependent only on the crystal space group and the unit cell dimensions. It is the relative intensities of the diffraction spots that are related to the electron density within the unit cell, i.e. the protein structure.

2.5.4 Reciprocal space

The reciprocal lattice is a formal construct developed by P. Ewald [220]. A set of lattice planes (hkl) are represented in reciprocal space (R^*) by a reciprocal lattice vector ($\mathbf{d}^*$), which is collinear with the interplanar distance vector $\mathbf{d}_{hkl}$ (Fig. 2.3b). The magnitude of $\mathbf{d}^*$ is equal to $1/d_{hkl}$ and therefore from Bragg's law we can state that $\mathbf{d}^* = 2\sin\theta/\lambda$ [215]. This diffraction condition can be visualised using the geometric construction known as the Ewald sphere [220, 215] (Fig. 2.3c). If the scattering diagram for Bragg's equation is placed into a sphere of radius $1/\lambda$ it becomes clear that $\mathbf{d}^* = |\mathbf{S}| = 2\sin\theta/\lambda$, where $|\mathbf{S}|$ is the magnitude of the scattering vector. When this relationship is satisfied then the diffraction condition is fulfilled and a reflection can be observed. Any reciprocal lattice point lying on the Ewald sphere will fulfil the diffraction condition. At any one

orientation of the crystal, only a small number of reciprocal lattice points will intersect the Ewald sphere. Rotation of the crystal during data collection will rotate the reciprocal lattice and change the points that are intersecting with the Ewald sphere, allowing more reflections to be recorded.

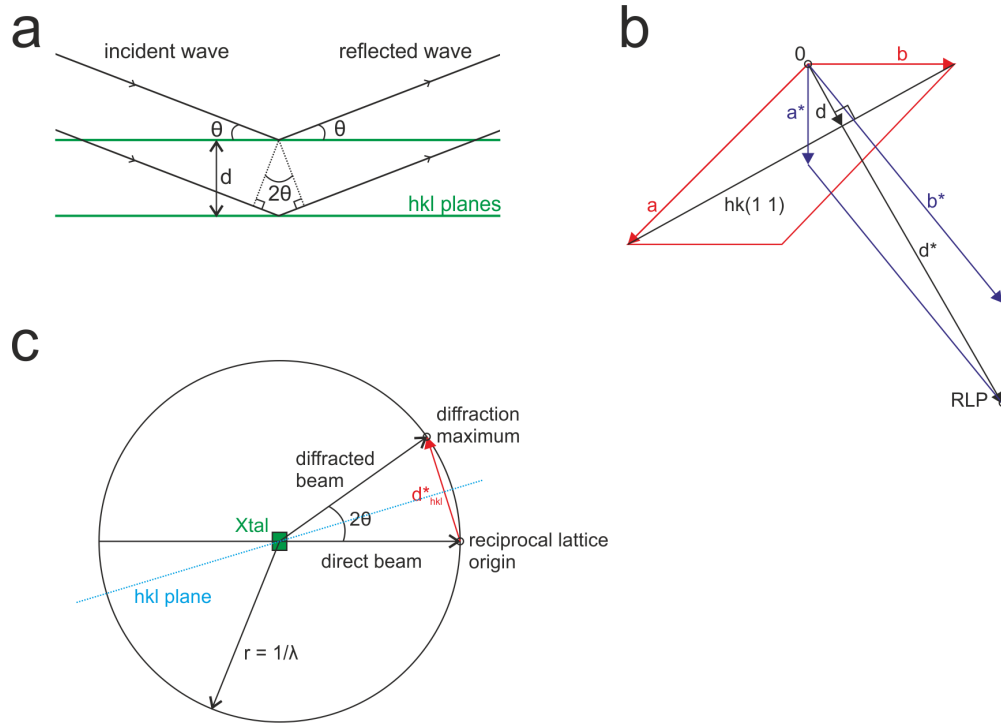


Figure 2.3: Bragg's law, the reciprocal lattice and the Ewald sphere construction. (a) The geometrical description of the Bragg equation. X-ray diffraction as reflection on hkl planes, separated by an interplanar distance, d is shown. The total path difference between the waves is $2d \sin \theta$. (b) A 2-dimensional reciprocal lattice construction. The real-space lattice is shown in red and the corresponding reciprocal lattice in blue. The reciprocal lattice vector $\mathbf{a}^*$ is perpendicular to $\mathbf{b}$, and $\mathbf{b}^*$ is perpendicular to $\mathbf{a}$. The hk plane (1 1) is shown. The reciprocal lattice vector, $\mathbf{d}^*$, extends from the reciprocal lattice origin (0) to the reciprocal lattice point (RLP) and is perpendicular to the hk plane and thus collinear with the interplanar distance vector, $\mathbf{d}$. (c) The Ewald sphere construction. The geometric description of Bragg's law is placed inside a sphere of radius $1/\lambda$ (λ is the incident beam wavelength) with the crystal (Xtal, shown in green) located at the centre of the sphere. The reflecting hk plane is shown in blue. The diffraction condition for this hk plane is fulfilled when its reciprocal lattice point lies on the Ewald sphere. The reciprocal lattice vector for this case is shown ($\mathbf{d}_{hkl}^*$, shown in red). Figures adapted from Figure 6-15 (a) and Figure 5-49 (b) and Figure 6-19 (c)[215].

2.5.5 Structure factors

From the Ewald sphere and Bragg's law we can identify where reflections will be collected but not the intensity of the reflection, as this is given by contents of the unit cell. In a diffraction experiment the intensities of all reflections are recorded and used to reconstruct the electron density within the unit cell. X-rays scattered by a protein crystal are the summation of a huge number of scattering events. The complex structure factor $\mathbf{F}_\mathbf{S}$ (Equation 2.2), a vector representing the diffracted X-rays, is a summation of the scattering contributions from every atom within the unit cell[215]. It depends explicitly on the scattering vector ($\mathbf{S}$) and the coordinate vectors ($\mathbf{r}_j$) of all atoms (j) in the crystal unit cell.

$$\mathbf{F}_\mathbf{S} = \sum_{j=1}^{atoms} f_{S,j}^0 \cdot \exp(2\pi i \mathbf{S} \mathbf{r}_j) \quad (2.2)$$

The atomic scattering factor $f_{S,j}$ is the reciprocal space representation of the atom's real-space electron density and can therefore be expressed as an integration of the electron density ($\rho_{at}(\mathbf{r})$) over the entire volume of the atom ($V(atom)$).

2.5.6 Reconstructing electron density

Similarly, the complex structure factor can be described as an integration of the electron density ($\rho(\mathbf{r})$) at any point ($\mathbf{r}$) in real-space over the entire volume of the unit cell ($V(cell)$) (Equation 2.3)[215], which is also equal to the Fourier transform of the real-space electron density.

$$\mathbf{F}(\mathbf{S}) = \int_r^{V(cell)} \rho(\mathbf{r}) \cdot \exp(2\pi i \mathbf{S} \mathbf{r}) d\mathbf{r} = FT[\rho(\mathbf{r})] \quad (2.3)$$

Logically an inverse Fourier transformation can be applied to this equation, from reciprocal space into real space, to give an equation for the electron density. The equation for electron density is best expressed as a discrete summation over the asymmetric unit (x, y, z) (Equation 2.4).

$$\rho(x, y, z) = \frac{1}{V} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} \mathbf{F}(hkl) \cdot \exp[-2\pi i(hx + ky + lz)] \quad (2.4)$$

The structure factor equation can be expressed as a similar summation (Equation 2.5).

$$\mathbf{F}(hkl) = V \sum_{x=0}^a \sum_{y=0}^b \sum_{z=0}^c \rho(x, y, z) \cdot \exp[2\pi i(hx + ky + lz)] \quad (2.5)$$

Using these two Fourier summations it is possible to easily convert between electron density and structure factors, a process that is important for structure solution. The complex structure factor $\mathbf{F}(hkl)$ comprises amplitude (F or $|\mathbf{F}|$) and phase (φ) terms. If this is incorporated into the electron density summation, for example at grid point x , it explicitly shows that both intensity and phase information is required for the reconstruction of electron density (Equation 2.6)[215].

$$\rho(x) = \frac{1}{V} \sum_{h=-\infty}^{+\infty} F_h \cdot \exp(i\varphi_h) \cdot \exp[-2\pi i(\mathbf{h}\mathbf{x})] \quad (2.6)$$

In a diffraction experiment only intensities are measured and these are proportional to the square of the structure factor amplitude. However, the critical phase information required for the reconstruction of the electron density is lost; this is known as the crystallographic phase problem. In order to reconstruct the electron density we must obtain the missing phases, there are two main strategies for this: experimental phasing (including single and multiple wavelength anomalous

diffraction) and molecular replacement. The technique employed throughout this thesis is molecular replacement and as such discussion of experimental phasing techniques are outside the scope of this text.

2.6 Structure solution pipeline

2.6.1 Data processing

Converting measured intensities or photon counts into structure factor amplitudes requires several corrections that are applied during raw data integration and scaling[221, 222]. The process of data integration involves three steps: autoint indexing, post-refinement and integration which comprises predicting the positions of Bragg reflections and obtaining an estimate of the intensity of each reflection[223]. Diffraction spots are essentially a distorted projection of the reciprocal lattice and using the Ewald construction it is possible to derive the reciprocal lattice vector of a reflection from its coordinates (Fig. 2.4). Spots lying on concentric lunes in the diffraction pattern correspond to reciprocal lattice points that intersected the Ewald sphere in the same plane, and these will have reciprocal lattice vectors of the same length (Fig. 2.4). Reciprocal lattice vectors collected for all spots on the diffraction image will therefore segregate into clusters, and the separation between clusters will relate to the vector between adjacent reciprocal lattice planes[223]. The Fourier transform of this will consist of peaks where the spacing corresponds to the real cell spacing; which can be used to define the dimensions of the unit cell[223]. The goodness of fit of these unit cell parameters to the tabulated possible expected dimensions is evaluated and the solution with the highest symmetry in the lowest penalty score group is selected as a working solution[224, 225]. In post-refinement the positions of the diffraction spots in X and Y and also over the rotation of the

crystal (ϕ) are used to refine parameters including the unit cell, detector position and orientation of the primary beam. Once accurate values for the unit cell and crystal orientation have been determined, reflection integration can occur. Spot positions can be predicted from the unit cell parameters. Reflection intensities are determined by three-dimensional profile fitting followed by subtraction of the background intensity, which is determined by pixel masking. This process is automated through various software including XDS[226], HKL2000[227] and MOSFLM[228]. Scaling, and an assessment of data quality follow the process of data integration.

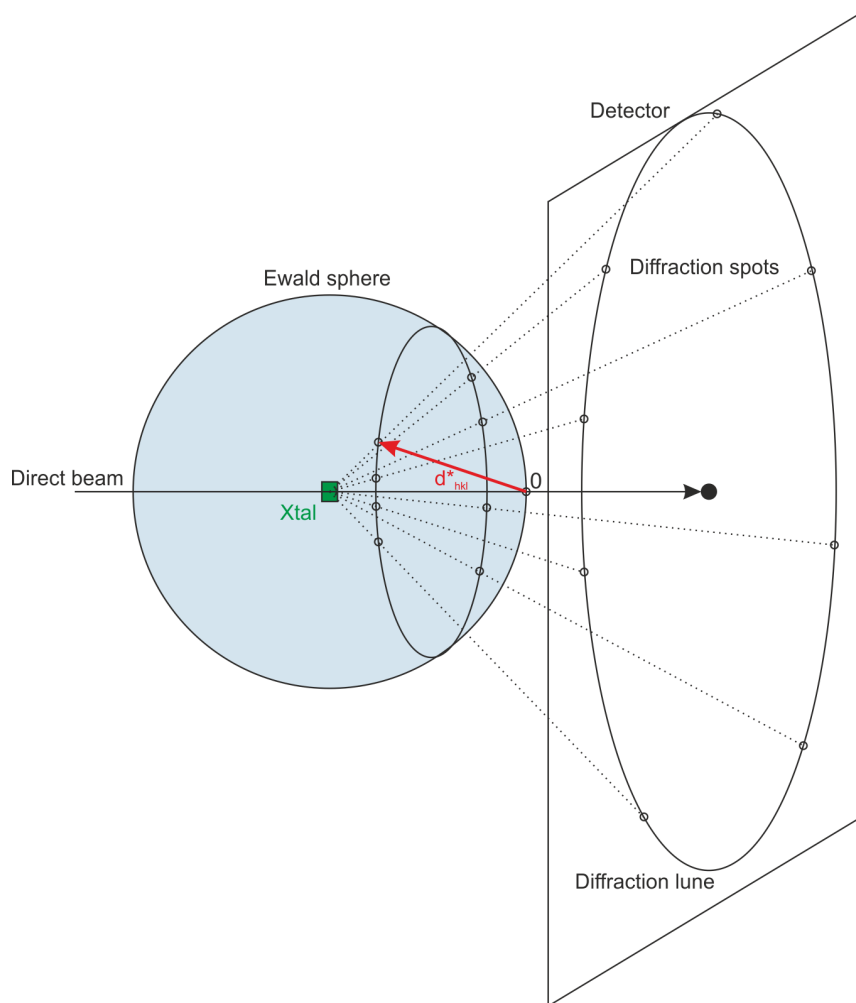


Figure 2.4: Relationship between the Ewald sphere and the diffraction pattern. The origin of the reciprocal lattice (0) is the point at which the direct beam exits the Ewald sphere. The crystal (Xtal, shown in green) is located at the centre of the Ewald sphere. The reciprocal lattice plane (1 0 0) is shown. Each diffraction spot on the detector can be mapped back to a scattering vector in reciprocal space (e.g. $\mathbf{d}^*_{hkl}$, shown in red). Figure adapted from[223].

Measured diffraction intensities are not all on the same scale because they are affected by parameters that vary throughout the experiment, including radiation damage and slow variation in beam intensity[229]. The redundancy of data collection, i.e. multiple measurements of symmetry-related reflections, is used to put all observations on a common scale. Following this, the internal consistency of the data can be assessed as a measure of data quality. Programs implementing this analysis include SCALA and XSCALE. X-ray data in this thesis were processed and scaled with the HKL suite[227] and XIA2[230], which implements

the programs XDS[226] and SCALA. Data collection and processing statistics for all of the structures solved in this thesis are shown in Table 1. The classical measure of data quality is the R_{merge} statistic (Equation 2.7), which measures the spread of ‘n’ independent measurements of the intensity of a reflection ($I_i(hkl)$) around their average ($\bar{I}(hkl)$)[231, 232].

$$R_{\text{merge}} = \frac{\sum_{hkl} \sum_{i=1}^n |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^n I_i(hkl)} \quad (2.7)$$

However, it was shown that the value of R_{merge} increases with data redundancy, despite the fact that a highly redundant data set should be of intrinsically higher quality than a data set in which each reflection is only measured once. This led to the introduction of the multiplicity-independent statistic R_{meas} , which adjusts the contribution of each reflection by a factor of $\sqrt{\frac{n}{n-1}}$ ([233]). Another data quality statistic is the precision-indicating merging R factor, $R_{\text{p.i.m.}}$ (Equation 2.8), which describes the precision of the averaged measurement.

$$R_{\text{p.i.m.}} = \frac{\sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{i=1}^n |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^n I_i(hkl)} \quad (2.8)$$

The value of $R_{\text{p.i.m.}}$ was shown to correlate well with improved refinement statistics and as such is a good predictor of the performance of a data set in structure determination[234].

The strength of diffraction decreases with resolution and so a high-resolution cut-off is applied to remove data with such weak intensities that their inclusion would reduce the quality of the resulting model. Historically, it was typical to truncate data at a resolution where the R_{merge} value is below 0.8 and the signal to noise ratio ($I/\sigma(I)$) is at least 2.0. However, recent studies have highlighted that this cut-off is too conservative and useful data that could contribute to

model quality are excluded by using these parameters. A better indicator of an appropriate high-resolution cut off is the $CC_{1/2}$ value, which calculates the correlation coefficient of the average intensities between random halves of a data set[232]. In test cases, including data to the resolution at which the $CC_{1/2}$ value was in the range 0.1-0.2 improved the quality of the final model, despite the traditional statistics greatly exceeded their conventional limits; the $I/\sigma(I)$ at these resolutions was between 0.3-0.6 and the R_{meas} values were over 300 % [235, 232]. Therefore, in addition to quoting both the R_{merge} and $I/\sigma(I)$ values, both the $CC_{1/2}$ and the $R_{\text{p.i.m.}}$ parameters are also displayed in Table 1 and were used to select the resolution cut-off [235, 236, 232, 234]. For all of the structures solved in this thesis, the $CC_{1/2}$ values remain >0.5 in the highest resolution shell, which indicates that these high-resolution reflections are contributing to the quality of the final model. This is despite very high R_{merge} (225 %) and low $I/\sigma(I)$ values (1.1) for some of the complexes (Table 1). The inclusion of these high-resolution data correlated with observed improvements in the electron density map, validating these criteria for establishing a data cut-off. This data processing strategy was made possible by the use of third generation synchrotron radiation sources (in this case the DIAMOND Light Source) with beamlines equipped with state-of-the-art detectors (in this case Pilatus 6D). This allows for the acquisition of excellent quality data, with high precision and accuracy, even for weaker reflections, due to the increased sensitivity and reduction of background scattering.

2.6.2 Molecular replacement

Following data processing, which yields structure factor amplitudes, the associated missing phases must be determined for reconstruction of the electron density. In this thesis phasing was carried out by molecular replacement. This

technique uses a model of known structure with a high similarity to the target structure ($>30\%$ sequence identity) to estimate phases allowing for structure solution. As the protein databank (<http://www.rcsb.org/pdb>) expands and the number of known structures increases, molecular replacement is likely to become an ever more important technique. The major task of molecular replacement is to conduct a rotational and translational search to place the model at the same orientation as the target within the unit cell[215]. This enables the model phases to be used for structure solution. Molecular replacement programs break this task into two distinct steps, the rotation function and the translation function. Traditionally molecular replacement programs used a Patterson-based approach to assess whether models were correctly placed[237]. The Patterson function assigns peaks to every interatomic vector and crucially the resulting map can be constructed from the intensities alone, without any phase information. This allowed Patterson maps for the model and the target to be compared, so directing the placement of the model. However, modern phasing software, such as PHASER[238], now implements maximum-likelihood approaches to molecular replacement, which have yielded improved success rates. The likelihood function describes the probability that an experimental value will be observed given the model. Generally, calculated structure factors (F_{calc}) from the model are compared to the observed structure factors from the data (F_{obs}). A ‘brute-force’ search procedure sampling angles (rotation function) and positions (translation function) of the model on a grid is performed, and the placement with the highest likelihood of matching the data is selected as the solution[238]. An unambiguous molecular replacement solution in PHASER is indicated by high log-likelihood gain (LLG) values of approximately 120. PHASER also outputs a Z-score which is a measure of the signal-to-noise of a search; it compares the LLG value to an LLG obtained from a set of a random rotations and translations. A Z-score of greater than 8 indicates an unambiguous solution, although Z-scores of less than 8 do not necessarily mean the correct solution has not been found.

In this thesis, the RGMC_{ND} -BMP2 complex was solved by molecular replacement in PHASER[238] using the structure of the disulphide bonded BMP2 dimer (PDB ID: 3BMP[168]) as a search model. Extra electron density accounting for two molecules of RGMC_{ND} in the asymmetric unit was immediately discernible after density modification in PARROT[239]. Density modification involves three main procedures; solvent flattening to remove noise, applying histogram matching of protein density to match the histogram of a well phased map and using non-crystallographic symmetry to enhance the signal to noise ratio. Following this, molecular replacement implemented in PHASER was used to solve the structures of the RGMA_{ND} -BMP2 and RGMB_{ND} -BMP2 complexes using the RGMC_{ND} -BMP2 complex as a search model; and the BMP2-eRGMB-NEO1_{FN56} complex using the RGMB_{ND} -BMP2 Form 1 (see Table 1) and the NEO1_{FN56}-RGMB (PDB ID: 4BQ6[240]) structures as search models. The eBMPr1A-BMP2 complex structure was also solved in a similar fashion, using the published complex structure as a search model (PDB ID: 1ES7[241]).

2.6.3 Model building and refinement

From the experimentally determined structure factor amplitudes and the phases supplied by molecular replacement an electron density map displaying the contents of the unit cells is produced. Due to limited resolution and imprecise phase information building and refining a protein model into the electron density is an iterative and meticulous process[242]. The first step is the construction of an initial model in the electron density map, which is often carried out using automated model building software such as BUCCHANEER[243]. The program begins by using a maximum likelihood method to locate $\text{C}\alpha$ positions in the electron density map by scoring possible positions and orientations in a six dimensional rotation and translation search. The highest scoring matches are then

used as ‘seed’ positions for polypeptide chain growth, which proceeds via a two-residue look-ahead approach[244]. This involves adding two new amino acid residues to the ‘seed’, subject to Ramachandran restraints[245]. The conformation of the polypeptide backbone can be described by three torsion angles for each amino acid, ϕ , ψ and ω . Due to the partial double bond character of the peptide bond ω is constrained to values of 0° for cis-peptides and 180° for trans-peptides. Only a subset of ϕ and ψ angle combinations are allowed for each amino acid and some combinations are particularly favoured; these values are mapped in the Ramachandran plot[245] and restrict placement of each residues during polypeptide chain growth. The best log-likelihood score for the placement of the new residues in the electron density map is calculated and used to select the position of the first residue. The second residue is then discarded having served the purpose of validating the position of the first. The result of chain growth is numerous overlapping fragments. From these a single coherent model is constructed by joining together consistent fragments and pruning inconsistent ones; of the different possible combinations the longest possible polypeptide trace is prioritised. The resulting trace is then analysed and edited manually in COOT[246], where small sections of the peptide chain can be refined to better fit the electron density.

From these initial estimated atomic positions better phases can be calculated, which allow the electron density map to be recalculated with a higher precision. From the new map, more accurate atomic positions can be determined, which in turn lead to further improved phases. In each cycle adjustments are made to the atomic model, and this eventually yields an accurate model of the target structure. This whole process is called structure refinement. Refinement programs such as autoBUSTER[247] and PHENIX[248] calculate a complete set of structure factor amplitudes from the atomic model and the calculated intensities are compared with the measured intensities and the model which best fits the

data is calculated using a maximum likelihood approach. A global indicator of model quality, the working R-factor (R_{factor}) (Equation 2.9), is used to assess how well the observed structure factors (F_O) correlate with the calculated model structure factors (F_C)[242].

$$R_{\text{factor}} = \frac{\sum ||F_O| - |F_C||}{\sum |F_O|} \quad (2.9)$$

However, R_{factor} is very susceptible to over-fitting as increasing the number of parameters used to describe the model can reduce its value. Therefore the R_{free} value was introduced[249] which is calculated in the same way as the R_{factor} but for a small subset of reflections that are not used in the model refinement. This therefore provides cross-validation, as it tests how well the model fits experimental data that was not used to fit the model. Values of approximately 25 % for the R-factors indicate that the model is a good fit to the data and the gap between R_{factor} and R_{free} should be minimised to prevent over-fitting. The quality of the model can also be assessed by analysing the fit of the model geometry to idealised values, for example the root mean squared deviation (r.m.s.d.) of bond angles and lengths from those of well-refined high-resolution structures and the Ramachandran statistics for each amino acid[242]. These statistics, along with the R_{factor} , R_{free} and the isotropic B-factors, are reported for each structure solved in this thesis as indicators of the quality of the refined model.

In the RGM_{CND}-BMP2 structure, the RGM_C polypeptide chain was traced using iterative rounds of the automated chain-tracing software BUCCANEER[243], manual model building in COOT[246] and refinement using autoBUSTER[247] and PHENIX[248]. This resulted in a well-defined model for the RGM_{CND}-BMP2 complex that included two molecules of RGM_C (residues Q36 to P129) bound to a disulphide linked BMP2 dimer (residues K293-R396). All complexes were refined using autoBUSTER[247] and PHENIX[248] and, where applicable,

non-crystallographic symmetry (NCS) restraints were applied[250]. Due to the dimeric nature of BMP2 the unit cells often contained more than one copy of each protein, which are related by NCS and the structures were restrained to be similar during the refinement. This increases the parameters used to describe each monomer and improves the accuracy of the model. In addition, for the BMP2-eRGMB-NEO1_{FN56} structure, target weight refinement as implemented in autoBUSTER[247] (using the individual high resolution structures of the BMP2, RGMB_{ND}, eRGMB (PDB ID: 4BQ6), NEO1_{FN5} and NEO1_{FN6} domains (PDB ID: 4BQ6) as targets) was applied[250]. Similarly to NCS restraints, this technique uses a similar previously solved structure of higher resolution to restrain the refinement. The stereochemical properties of all of the models for the complexes solved in this thesis were assessed using the webserver MOLPROBITY[251]. Crystallographic and Ramachandran statistics for all structures solved in this thesis are given in Table 1. Superpositions displayed throughout this work were calculated using the program COOT[246] and electrostatic potentials were generated using APBS[252] as implemented in PYMOL[253]. Buried surface areas of protein-protein interactions were calculated using the PISA webserver[254] for a probe radius of 1.4 Å.

	RGMA_{ND}- BMP2 complex	RGMB_{ND}- BMP2 complex Form 1	RGMB_{ND}- BMP2 complex Form 2	RGMC_{ND}- BMP2 complex	eBMPR1A- BMP2 complex	eRGMB- BMP2- NEO1_{FN56} complex
DATA COLLECTION						
X-ray source	DIAMOND- I04	DIAMOND- I03	DIAMOND- I04-1	DIAMOND- I03	DIAMOND- I03	DIAMOND- I03
Resolution	45.0-3.2 (3.42-3.20)	50.0-2.85 (2.93-2.85)	45.0-2.7 (2.80-2.70)	50.0-2.35 (2.43-2.35)	26.3-2.6 (2.67-2.60)	103.5-3.2 (3.18-3.10)
Space group	P3 ₂ 21	I2 ₁ 3	P3 ₂ 21	P2 ₁ 2 ₁ 2 ₁	P2 ₁	P4 ₃ 2 ₁ 2
Cell dimensions [Å]	a = 83.9 Å b = 83.9 Å c = 114.8 Å	a = 129.0 Å b = 129.0 Å c = 129.0 Å	a = 85.0 Å b = 85.0 Å c = 115.2 Å	a = 69.0 Å b = 76.3 Å c = 81.7 Å	a = 66.4 Å b = 75.8 Å c = 115.0 Å	a = 120.1 Å b = 120.1 Å c = 204.1 Å
Unique reflections	8086 (1440)	8429 (626)	13647 (1361)	18558 (1827)	34888 (2432)	27477 (1933)
Completeness [%]	100 (100)	99.4 (98.9)	99.6 (99.4)	99.9 (100.0)	99.0 (93.5)	98.5 (95.5)
R _{merge} [%]	14.7 (74.1)	4.8 (81.7)	10.1 (59.3)	13.9 (79.0)	8.5 (93.5)	8.2 (225.1)
R _{p.i.m.} [%]	6.0 (29.2)	2.8 (50.3)	3.9 (38.8)	5.4 (71.2)	6.7 (71.6)	3.1 (83.8)
CC _{1/2}	0.994 (0.970)	0.998 (0.550)	0.999 (0.934)	0.998 (0.589)	0.999 (0.817)	0.999 (0.749)
I/σI	9.1 (3.0)	15.1 (1.5)	14.7 (2.7)	11.7 (2.0)	9.6 (1.3)	19.0 (1.1)
Redundancy	7.1 (7.3)	4.2 (4.0)	9.6 (9.0)	5.7 (5.1)	3.0 (2.7)	8.8 (8.9)
REFINEMENT						
Resolution range [Å]	45.0-3.2 (3.58-3.20)	50.0-2.85 (3.19-2.85)	45.0-2.8 (3.07-2.80)	45.0-2.35 (2.49-2.35)	26.3-2.6 (2.68-2.60)	50.0-3.1 (3.18-3.10)
No. of reflections	8074 (2156)	8427 (2347)	12284 (2884)	18135 (2495)	34858 (2785)	26105 (1837)
No. of atoms (protein/sulphate/citrate/ chloride/MPD/nitrate/ EG/sugar/H2O/tartrate/ acetate)	2112/0/0/ 0/0/0/ 0/0/0/0/ 0	1343/40/0/ 0/0/0/ 0/0/0/0/ 0	2017/0/13/ 1/8/0/ 0/0/20/0/ 0	273/0/0/ 2/0/24/ 3/0/85/0/ 0	6033/0/0/ 6/0/0/ 12/233/83/0/ 0	4231/0/0/ 0/0/0/ 0/39/0/30/ 4
B factors [Å] (protein/sulphate/citrate/ chloride/MPD/nitrate/ EG/sugar/H2O/tartrate/ acetate)	110/0/0/ 0/0/0/ 0/0/0/0/ 0	110/180/0/ 0/0/0/ 0/0/0/0/ 0	88/0/80/ 107/72/0/ 0/0/88/0/ 0	55/0/0/ 66/0/76/ 61/0/47/0/ 0	70/0/0/ 6/0/0/ 12/104/83/0/ 0	94/0/0/ 0/0/0/ 0/185/0/158/ 106
R _{factor} [%]	22.7 (23.3)	21.7 (24.5)	23.3 (34.5)	18.7 (20.8)	19.1 (26.1)	20.2 (44.0)
R _{free} [%]	26.4 (25.4)	23.6 (27.6)	26.6 (42.3)	23.8 (24.5)	21.9 (28.4)	23.3 (36.9)
r.m.s.d. bonds [Å]	0.010	0.008	0.010	0.010	0.008	0.007
r.m.s.d. angles [Å]	1.15	0.95	1.07	1.09	1.04	1.25
Ramachandran statistics						
Favoured [%]	92.1	95.9	95.2	95.6	97.0	92.8
Disallowed [%]	0.8	0	0	0.9	0.1	0.4

Table 1: Crystallographic data collection and refinement statistics. Each structure was determined from one crystal. Numbers in parentheses refer to the highest resolution shell. R_{free} = the R-factor against 5 % of the data, r.m.s.d. = root mean square deviation from ideal geometry, R_{p.i.m.} = precision indicating merging R-factor[234], CC_{1/2} = correlation coefficient between random half data sets[236, 232].

2.7 Surface Plasmon Resonance (SPR)

SPR is an optical technique that can be used to measure protein interactions. One protein, known as the ligand, is immobilised onto the surface of sensor chip, which consists of four flow cells. Other protein binding partners, known as analytes, are flowed over the chip in an aqueous buffer. The sensor chip consists of a glass surface coated in a layer of gold followed by a carboxymethylated dextran matrix upon which proteins are immobilised. In the instrument, beams of light (670 nm wavelength) exceeding the critical angle are totally internally reflected from the glass surface of the sensor chip. During this process some energy is lost to the gold particles coating the glass, through interaction of the incident light wave with the surface plasmons (free electrons) on the gold surface. Maximum energy is lost to the gold film at a particular incident angle (θ_{spr}), which is measured using a photo-detector array to record the angle at which the intensity of the reflected light is at a minimum. The value of θ_{spr} depends on the refractive index of the medium directly adjacent to the gold surface, which is affected by the binding of the analyte to the immobilised ligand. Therefore, the binding of any proteins to the surface of the sensor chip is detected as a change in θ_{spr} and these changes are reported as response units (RU)[255].

SPR experiments were performed using a BIAcore T200 machine (GE Healthcare) at 25 °C in SPR running buffer (10 mM HEPES, pH 7.5, 150 mM NaCl, 0.005 % (v/v) polysorbate 20). The surface of the CM5 biosensor chip was activated using N-hydroxysuccinimide and ethyl(dimethylaminopropyl) carbodiimide (NHS/EDC) and following protein or streptavidin coupling, blocked using ethanolamine. All experiments were performed using direct protein immobilisation by amine coupling to CM5 biosensor chips except for those where biotinylated eRGMB was immobilised onto streptavidin-coupled CM5 biosensor chips (as specified in the figure legend)[256]. Analytes were dialysed against SPR

running buffer prior to use, and 1 in 2 dilution series were prepared for analysis. Typically BMP2 was bound to the surface at concentrations of 100-1000 response units. For the eBMPR1A-eRGMB competition experiment a stock solution of 20 μ M eBMPR1A and 2.5 μ M eRGMB was dialysed against SPR running buffer prior to use and 1 in 2 dilution series were prepared in buffer containing 2.5 μ M eRGMB, resulting in a dilution series of eBMPR1A in a constant (2.5 μ M) concentration of eRGMB. BMP2 was bound to the surface at concentrations of 150, 500 and 1000 response units. After each binding experiment where BMP2 was used as ligand, the chip was regenerated by 120s at 20 μ l/min bursts of 4 M Urea, 50 mM Tris pH 8.0, 150 mM NaCl. All experiments were done in duplicate with independently purified proteins. The signal from experimental flow cells was processed, and corrected by subtraction of the nearest blank measurement (every third injection was a blank) and the reference signal from a blank flow cell. In all experiments, the experimental trace returned to baseline after each regeneration. All data were analysed using SCRUBBER2 (Biologic) and GraphPad Prism Version 6.04 (GraphPad Software, La Jolla California USA). Best-fit binding curves were calculated for BMPR1A-BMP and BMPR2-BMP interactions using non-linear curve fitting of a one-site - specific binding model ($Y = B_{\max} * X / (K_d + X)$), where X is the analyte concentration and $B_{\max}$ is the maximum analyte binding). Best-fit binding curves were calculated for BMP2-RGM interactions using non-linear curve fitting of a one-site - total binding model ($Y = B_{\max} * X / (K_d + X) + NS * X + \text{Background}$, where X is analyte concentration and the amount of non-specific binding is assumed to be proportional to the concentration of analyte, hence NS is the slope of non-specific binding). The background value was constrained to zero as the data had been previously referenced. $B_{\max}$ and K_d values reported are determined for the specific binding component only. A best-fit binding curve was calculated for the eBMPR1A-eRGMB competition experiment using a log(agonist) vs response - variable slope model ($Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + 10^{((\text{LogEC50} - X) * \text{HillSlope}))}$), where X is analyte concentration,

top and bottom are plateaus in the units of the Y axis and EC50 is the concentration of agonist that gives a response half way between bottom and top). The HillSlope parameter was constrained to 1.0. An R^2 value to quantify goodness of fit (range of 0-1.0) was reported. An IC_{50} value of $4.0 \pm 1.0 \mu M$ was calculated. Given the concentration of eRGMB was $2.5 \mu M$, 1.7 fold molar excess of eBMPr1A is required to effectively compete with eRGMB in solution.

2.8 Small-angle X-ray scattering (SAXS)

2.8.1 Basic principles of SAXS

SAXS is an important complementary technique in structural studies. It can provide information about protein complex stoichiometry, conformation, and flexibility in the relatively low-resolution range of 10-50 Å, which crucially for this study, can help to confirm the solution nature of protein complex assemblies observed in crystal structures. Protein molecules in solution scatter X-rays to produce a diffuse scattering pattern and the scattering curve for a protein ($I(q)$) is generated after subtraction of the scattering from a buffer blank. The scattering curve $I(q)$ is a function of the momentum transfer (q), which is a measure of the directional momentum change that the photons undergo and is dependent on the scattering angle (2θ) and the wavelength of the incident X-ray beam (λ) (Equation 2.10)[257].

$$q = \frac{4\pi \sin \theta}{\lambda} \quad (2.10)$$

At low resolution the scattering can be described by the Guinier approximation (Equation 2.11)[257].

$$I(q) = I(0) \exp \frac{-(q^2 R_g^2)}{3} \quad (2.11)$$

The Guinier Plot of $\log(I(q))$ against q^2 will give a straight line from which the parameters R_g and $I(0)$ can be derived. The parameter R_g is the radius of gyration, which describes the average squared distance of each scatter from the centre of the scattering particle. The other parameter $I(0)$ is the intensity measured at $q = 0$, which can be extrapolated from the Guinier plot, and corresponds to the square of the number of electrons in the scattering particle and therefore can be useful for molecular weight determination. However, calculation of the molecular weight from $I(0)$ requires calibration with a reference sample such as BSA, and so a true reference-free method for molecular weight determination is implemented in this thesis involving the calculation of the parameter V_c (Equation 2.12)[258].

$$V_c = \frac{I(0)}{\int q I(q) dq} \quad (2.12)$$

V_c is the ratio of the particle's $I(0)$ to its total scattered intensity, which is the integrated area of the curve $qI(q)$ against q ; and, importantly, V_c can be directly calculated from the scattering curve. A log-log plot of both (V_c^2/R_g) against molecular weight is linear, allowing the molecular weight of the scattering particle to be determined from these two parameters. Both V_c and R_g are contrast and concentration independent and therefore yield a more accurate estimate of molecular weight than $I(0)$.

A Fourier transform of the scattering curve $I(q)$, known as the pair-distribution function ($P(r)$), provides real-space information about the distances between electrons within the scattering protein molecules (Equation 2.13)[257].

$$P(r) = \frac{r}{2\pi^2} \int_0^\infty I(q) q \sin(qr) dq \quad (2.13)$$

The maximum linear dimension of the scattering particle ($D_{\max}$) can be extracted from the pair-distribution function, as the $P(r)$ function is equal to zero at $r = 0$ and $r \geq D_{\max}$. Importantly, $P(r)$ functions can be readily calculated from high-resolution atomic models and an inverse Fourier transform will then yield a corresponding theoretical scattering curve. These can be compared to experimental scattering curves to determine how well a crystal structure represents the solution state of a protein or protein complex. The agreement between two scattering curves can be assessed using the χ^2 metric (Equation 2.14), where K is the number of experimental points, μ is a scaling factor and $\sigma(q_j)$ is the standard deviation[259].

$$\chi^2 = \frac{1}{K-1} \sum_{j=1}^K \left[\frac{\mu I(q_j) - I(q_j)_{\text{experiment}}}{\sigma(q_j)} \right]^2 \quad (2.14)$$

However, caution should be employed when interpreting χ^2 values as this measure weights the lower resolution data more strongly, and fitting to measurements with high errors can still yield good χ^2 values.

2.8.2 SAXS data collection and processing

Data were collected at beamline BM29 at the ESRF (Grenoble, France)[260] at 293 K within a momentum transfer range of $0.01 \text{ \AA}^{-1} < q < 0.45 \text{ \AA}^{-1}$; giving a maximum nominal resolution of $14 \text{ \AA}$. X-ray wavelength was $0.995 \text{ \AA}$ and data were collected on a Pilatus 1M detector. Protein samples were measured at the following concentrations: NEO1_{FN56}: 1.06 and 4.94 mg/ml; eRGMB: 0.96 and 5.48 mg/ml; NEO1_{FN56}-eRGMB: 0.51 and 1.01 mg/ml; BMP2-eRGMB-

NEO1_{FN56}: 0.68 and 1.14 mg/ml. Data reduction and calculation of invariants was carried out using standard protocols implemented in the ATSAS software suite[261]. A merged dataset was obtained by combining the low-angle part of the low-concentration dataset with the high-angle part of the high-concentration dataset. Molecular weight determination was performed using the volume of correlation metric V_c as implemented in the program Scatter[258]. The BMP2-eRGMB-NEO1_{FN56} solution structure was modelled starting from the crystal structure of the ternary complex. C-terminal eRGMB residues His335-Ser410 were not observed in the electron density; a structural model for the domain was calculated using the program ROSETTA3.5[262], and constrained by imposing alpha-helical secondary structure and a disulphide bond between Cys358 and Cys372. Missing loops and N- and C-termini were added in extended conformations using the program Modeller[263]. Starting models for NEO1_{FN56}, eRGMB, and the eRGMB-NEO1_{FN56} binary subcomplex were then extracted from the completed BMP2-eRGMB-NEO1_{FN56} model. All-atom ensemble modelling of NEO1_{FN56}, eRGMB, eRGMB-NEO1_{FN56} and BMP2-eRGMB-NEO1_{FN56} was performed using the program AllosMod[264]; in each case, 50 independent ensembles of 100 models were generated. From this pool, automated selection of the minimal set of models satisfying the scattering data was performed using the program MES[259], and calculation and fitting of scattering patterns was performed using the program FoXS[265]; this whole procedure was automated using the AllosMod-FoXS web server[266]. The solution structures of NEO1_{FN56}, eRGMB, eRGMB-NEO1_{FN56} and BMP2-eRGMB-NEO1_{FN56} are described by one, one, three and two models, respectively.

2.9 Multiangle light scattering

MALS measures the scattering of laser light by a solution of macromolecules. When connected inline with size exclusion chromatography the precise molecular weights of protein peaks can be determined, giving an indication of the complex species present in solution. MALS experiments were carried out using a DAWN HELEOS II (Wyatt Technology), equipped with a K5 flow cell and a 30 mW linearly polarised GaAs laser with a wavelength of 690 nm. Proteins used for MALS contained wild-type sugars. Prior to the experiments, individual proteins were purified by SEC. The BMP2-RGMB_{ND} and BMP2-eBMPRI1A complexes were formed by mixing the individual protein components in a 1:1 molar ratio. Complexes were dialysed against buffers of either pH 7.5, pH 6.5 or pH 5.5 (generated using the MMT buffer system (Malic acid, MES, Tris; Molecular Dimensions); 10 mM MMT, pH 5.5/6.5/7.5, 150 mM NaCl). Complexes were concentrated to 2 mg/ml prior to MALS analysis. Data were analysed using ASTRA (Wyatt Technologies) and molecular weights were calculated using the Debye fit method. Graphs were produced using GraphPad Prism Version 6.04 (GraphPad Software, La Jolla California USA).

2.10 Luciferase reporter assay

This assay made use of the BMP-responsive element luciferase reporter plasmid BRE-Luc, which contains two repeats of the minimal elements comprising the Id1 promoter, a specific target of downstream BMP signalling. The minimal regions were identified using site-directed mutagenesis methods, and include two SMAD binding elements: the GC-palindromic sequence GGCGCC and CGCC sequences flanked by two CAGC motifs. In luciferase reporter experiments, activation of the promoter results in expression of the enzyme luciferase, which converts the

substrate luciferin to oxyluciferin in a reaction that produces light. The amount of light produced is proportional to the level of activation of the promoter and therefore using BRE-Luc, changes in the relative light units recorded indicate changes in the level of BMP signalling[267].

LLC-PK1 cells were plated in complete DMEM supplemented with 10 % FBS at a density of 5×10^4 cells per well in a 96-well plate (Nunc-Immuno™ MicroWell™ 96 well polystyrene plates, Sigma-Aldrich). After 24 h LLC-PK1 cells were transfected using Lipofectamine 2000 transfection reagent (Lifetechnologies) according to manufacturer's protocol, with 40 ng pGL3 BRE-Luciferase plasmid[267], 30 ng Renilla control plasmid and, where indicated, 20 ng of either empty pHLSec vector control or full-length RGMB. Eight hours post-transfection the cells were washed with PBS (100 μ l) and serum starved in complete DMEM supplemented with 0.1 % FBS overnight. For membrane-attached RGM constructs, cells were stimulated with 6 nM BMP2 or buffer. To test the effects of soluble constructs, cells were stimulated with 25 nM BMP2 pre-incubated with a dilution series (from 2.5 μ M to 10 nM) of eBMPr1A, eACVR2A and eRGMB. After 48 hours incubation, cells were washed with PBS, lysed, and luciferase activity measured using a dual luciferase assay system (Promega) according to the manufacturer's instructions. Luminescence was quantified using a luminometer (Tecan, Infinite 200 PRO). Graphs were produced using GraphPad Prism Version 6.04 (GraphPad Software, La Jolla California USA). Multiple comparisons were assessed with the one-way ANOVA and Bonferroni post-testing (GraphPad Software; Prism).

2.11 Fluorescence microscopy

Fluorescence microscopy allows tracking of specific fluorescently labelled proteins within living cells over time. However, the technique has traditionally been lim-

ited in resolution to approximately half the wavelength of visible light (~ 200 nm), which does not permit observation of protein assemblies. In addition background fluorescence from proteins localised to the cell interior has limited observation of protein signalling assemblies in the membrane. Total internal reflection fluorescence (TIRF), direct Stochastic Reconstruction Microscopy (dSTORM) imaging represents one method of overcoming these issues to allow observation of membrane protein clusters in live and fixed cells. Additionally, protein-protein interactions can be analysed in live cells using fluorescence resonance energy transfer (FRET), recorded by fluorescence lifetime imaging (FLIM).

2.11.1 Principles of TIRF-dSTORM

TIRF microscopy involves the use of fluorescence light waves that are totally internally reflected at the glass-water interface between the specimen and the coverslip resulting the production of an evanescent wave, which decays exponentially from this interface. Only fluorophores within approximately 100 nm from the interface are excited, which limits the study to the plasma membrane and the cytosolic region directly adjacent. This reduction in background fluorescence greatly enhances the signal from membrane-localised proteins and enables the study of signalling events.

Traditionally the resolution limit of fluorescence light microscopy is described by Abbe's formula (Equation 2.15) and is dependent on the wavelength of the fluorescent light (λ) and the numerical aperture (NA) of the objective.

$$Resolution_{x,y} = \frac{\lambda}{2NA} \quad (2.15)$$

This is normally roughly equal to half the wavelength of the fluorescence light

and the resolving power of fluorescence microscopy is generally considered to be limited to ~ 200 nm. However, recently, advanced super-resolution microscopy techniques including dSTORM, have broken this diffraction limit and achieved resolutions of close to 20 nm. These techniques determine the precise localisation of each fluorophore by recording the fluorescence emission pattern of photons (point spread function (PSF)) on a detector and averaging this as a two-dimensional Gaussian function with the fluorophore localised to the peak of the function[268, 269]. The localisation precision depends on the number of photons collected and the standard deviation of the PSF. The technique depends on recording only a sub-set of the fluorophores at any one time to ensure that the PSFs do not overlap, and this is achieved by forcing the majority of fluorophores into a recoverable dark-state using high powered excitation lasers[270, 268, 269].

By estimating the precise position of each fluorescently labelled protein, the size, shape and number of clusters on the membrane can be estimated. Changes in this clustering level over time, in response to stimuli can also be measured using fixed cell samples or using a time-gated approach in live cells; both methods were implemented in this thesis.

2.11.2 Principles of FRET-FLIM

Fluorescence resonance energy transfer (FRET) is a non-radiative (long-range dipole-dipole interaction) process by which energy is transferred from the excited state of a donor fluorophore to an acceptor fluorophore. The FRET efficiency represents the probability that the excited donor will return to its ground state by FRET and is dependent on:

1. the distance between the donor and acceptor fluorophores (R).
2. the overlap between donor emission and acceptor excitation spectra ($J(\nu)$).

3. the quantum yield of the donor (Φ_D).
4. the fluorescence lifetime of the donor.
5. the relative orientations of the donor and acceptor dipoles.

The FRET efficiency (E) is greatly dependent on distance between the fluorophores (Equation 2.16).

$$E = \frac{R_0^6}{R_0^6 + R^6} \quad (2.16)$$

Where R_0 (Å) is the Förster distance (the distance between the fluorophores at which $E = 50\%$). R_0 can be calculated from a number of spectroscopic and dipole-related parameters[271] (Equation 2.17).

$$R_0 = \frac{9000(\ln 10)\Phi_D\kappa^2J(v)^{\frac{1}{6}}}{128\pi^5N_A n^4} \quad (2.17)$$

Where N_A is Avagadro's number, n is the refractive index of the medium and κ is the orientation factor (which depends on the angles between the donor and the acceptor dipoles). R_0 generally ranges between 2-8 nm and therefore a measurable FRET efficiency between a donor-tagged protein and an acceptor-tagged protein is interpreted as a protein-protein interaction[271].

The lifetime of the donor fluorescence depends on the rate of the decay of the excited state. When FRET is occurring the fluorescence lifetime of the donor decreases relative to the lifetime measured in the absence of acceptor. The donor mTFP is commonly used as this fluorophore exhibits a mono-exponential lifetime decay. In the presence of acceptor some proportion of the donor population will be bound to acceptor resulting in a bi-exponential fluorescence decay with two lifetimes; the FRET lifetime of donor bound to acceptor (τ_{DA}) and the lifetime

of the donor alone (τ_D). Analysis of FLIM data will recover these two values, allowing both the FRET efficiency and the fraction of interacting donor (f_D) to be determined. The spatiotemporal evolution of the interaction can then be tracked in live cells[271].

2.11.3 Sample preparation

BMP2 was stored at 4 °C in 10 mM ammonium acetate buffer. Prior to imaging a 40 nM BMP2 solution was dialyzed against phenol red-free complete DMEM supplemented with 0.1 % FBS, to remove excess ammonium acetate to prevent buffer effects. For imaging, COS7 cells were seeded at a density of 1×10^5 cells/well in glass bottom dishes (MatTek). After 24 hours, cells were transfected with constructs for expression of fluorescently-tagged proteins using Lipofectamine 2000 transfection reagent (Lifetechnologies) according to the manufacturer's protocol. Cells were transfected with 1.5 μ g NEO1-mVenus + 1.5 μ g full-length RGMB for dSTORM and with either 1.5 μ g BMPR2-mTFP or 1.5 μ g BMPR2-mTFP + 1.5 μ g BMPR1A-mVenus for FRET-FLIM. Eight hours post-transfection the cells were serum-starved in phenol red-free complete DMEM supplemented with 0.1% FBS overnight. Prior to imaging the media was removed from the cells and either BMP2 containing or buffer blank media was added. To generate fixed samples for dSTORM, COS7 cells were incubated with either BMP2 containing or buffer blank media for the indicated amount of time. The media was then removed and cells were fixed in 4 % paraformaldehyde (PFA) for 10 min at room temperature. Fixed cells were washed thoroughly in PBS to remove excess PFA and stored in PBS at 4 °C.

2.11.4 dSTORM imaging

dSTORM images were acquired on a TIRF microscope (ELYRA; Zeiss) with a 100 oil-immersion objective with a numerical aperture of 1.46. For illumination/photoconversion, 30 % transmission of the 488-nm laser and 1 % transmission of the 405-nm laser were used and 5,000-10,000 images were acquired per sample with a cooled, electron-multiplying charge-coupled device camera (iXon DU-897D; Andor). Exposure time was 300 ms. Recorded images were analysed with Zeiss ZEN software.

2.11.5 dSTORM data processing

Raw fluorescence intensity images were analysed using the software Zen 2010D (Zeiss MicroImaging). A Gaussian and Laplace filter was applied to each frame, and overlapping events were excluded. An event was classified as originating from a single molecule when $I - M > 6S$ (where I is event intensity, M is mean image intensity and S is the s.d. of image intensity). The centre of each point-spread function was then calculated by fitting to a two-dimensional Gaussian distribution and a table containing the x-y particle coordinates of each molecule was extracted. Non-overlapping regions of interest, $5\ \mu\text{m} \times 5\ \mu\text{m}$ or $3\ \mu\text{m} \times 3\ \mu\text{m}$ in area, containing the two-dimensional molecular coordinates were cropped for analysis using Origin (Northhampton, MA, USA), and events with localization precision worse than 60 nm were discarded. Areas containing ~ 300 molecules for live cell imaging and ~ 1000 events for the fixed samples were selected. To analyse the spatial point pattern, we used the Ripley's K-function, calculated with SpPack[272] and the Spatstat package for R software[273], and plotted the L-function as described[268, 269]. Briefly, Ripley's K-function is a measure of the number of points encircled by concentric circles of radius r centred on each point. K-values from the Ripley function therefore scale with circle area and

so are transformed into the L-function. Using this equation, scaling is linear with the radius. Random distributions have an $L(r)$ value of r over all r values. Therefore to analyse levels of clustering we plotted $L(r)$ against r ; and positive values at a given r indicate clustering at that spatial scale. Quantitative cluster maps were generated using Getis and Franklin's analysis as described[268, 269]. Briefly, $L(r)$ values at a spatial scale of 50 nm ($L(50)$) were computed for each point with R and interpolated using Origin to produce a quantitative cluster map. This was then pseudo-coloured with ImageJ[274] to highlight regions of high clustering. Live cell dSTORM images were constructed using a time-gated window approach[268, 275]. On the basis of the frame number from the raw data acquisition, cluster maps were generated from 5000 continuous frames, with each cluster map shifted by 2000 frames relative to the previous one. This corresponded to a time-gated window of ~ 7 minutes with the clustering analysed over a total of 30 minutes.

2.11.6 FRET-FLIM imaging

Imaging was performed using a Leica TCS SP8 confocal laser scanning microscope equipped with a white light laser. We used excitation lines 470 nm and 514 nm to excite mTFP and mVenus respectively. For lifetime fluorescence resonance energy transfer (FRET-FLIM) imaging the system was equipped with a Pico-harp 300 time-correlated single photon counting (TCSPC) board (Picoquant, Berlin, Germany). A 440 nm PDL 800-B pulsed diode laser was used to excite mTFP (40 MHz) and detected using a CFP (460/500) emission filter cube and Leica HyD detector. Time-resolved data, acquired through an acquisition period during which 500 photon counts per image were collected, were analysed using Picoquant SymPhoTime 64 software. Fluorescence lifetime was calculated using a bi-exponential tail-fit to the TCSPC histogram and applied to each pixel in

the image with a maximum likelihood estimator to produce the final fluorescence lifetime image.

2.11.7 FRET-FLIM Image Analysis

The pixel-by-pixel fluorescence lifetime and fluorescence intensity information were exported as separate ASCII files from SymPhoTime 64 (PicoQuant) into ImageJ[274]. The Robust Automatic Threshold Selection Plugin (B. Tupper & M. Sieracki, ImageJ[274]) was applied to the intensity image to create a mask of regions of high fluorophore intensity[276]. This mask was then applied to the fluorescence lifetime image to isolate the pixel lifetimes in these areas. The average lifetime of the pixels within the mask was calculated from a histogram of the pixel lifetimes. Mean lifetime and $f(D)$ values were calculated from the double exponential fit to the TCSPC histogram[277]. Graphs were produced using Prism and figures were produced using ImageJ and CorelDRAW X6 (COREL).

2.11.8 Live-cell imaging of fluorescently-labelled BMP2

BMP2 was mixed in a 1:1 ratio with RGMB_{ND} to form the BMP2-RGMB_{ND} complex, which was dialysed into PBS and concentrated to 2 mg/ml. Stock borate buffer (40 μ l, 0.67 M, Thermo Scientific DyLight 633 Antibody Labeling Kit) was added to 0.5 ml of the complex for a final concentration of 50 mM sodium borate pH 8.5. The protein solution was added directly to the vial of NHS-activated DyLight 633 (Thermo Scientific DyLight 633 Antibody Labeling Kit), mixed, and incubated in the dark at room temperature for 1 hour. The labelled complex was then dialysed twice against 1 L 4 M Urea, 100 mM Tris-HCl, pH 6.0, 5 mM EDTA (4 °C, 24 h) and the labelled BMP2 and RGMB_{ND} were separated by size exclusion in this buffer. Fractions containing BMP2-DyLight 633, identified by

their blue colour and confirmed by SDS-PAGE, were pooled for use in live-cell imaging. BMP2-DyLight 633 was dialysed twice against 1 x PBS (1 L, 4 °C, 24 h). The activity of labelled BMP2-DyLight 633 was confirmed by luciferase assay (see above). C2C12 cells were seeded at a density of 1×10^5 cells/well in glass bottom dishes (MatTek). After 24 hours, cells were transfected with Rab5a-mCherry (an early endosome marker) using Lipofectamine 2000 transfection reagent (Lifetechnologies) according to the manufacturer's protocol. Eight hours post-transfection the cells were serum-starved in complete DMEM supplemented with 0.1 % FBS overnight. Prior to imaging the cell media was exchanged for 1 ml of 1 x PBS and subsequently 20 μ l BMP2-DyLight was added to the cells. Imaging was performed using a Leica TCS SP8 confocal laser scanning microscope equipped with a white light laser using excitation lines of 561 nm (2 %) and 634 nm (20 %). Fluorescence emission was detected using a Leica HyD detector between wavelengths 595-615 nm for mCherry and 654-680 nm for Dylight633. Imaging speed = 400 Hz, Z-stack images were collected, one image every 15 seconds, over a 30 minute time series.

2.12 Analytical ultracentrifugation

Sedimentation velocity (SV) AUC experiments are ideal for analysing dynamic and low affinity protein-protein interactions. The nature of migration of macromolecular assemblies within a gravitational field means that faster sedimenting protein complexes can be consistently assembled from a slower sedimenting pool of individual components. The migrations can be observed in a way that reflects the equilibrium and kinetic properties of the interaction[278]. Furthermore, much higher protein concentrations can be sampled than is possible, for example, in a MALS experiment.

SV-AUC experiments were performed using a Beckman Optima XL-I analytical centrifuge and a run temperature of 20 °C. After gel filtration protein samples were concentrated to the following concentrations for SV analysis: eRGMB^{WT} and eRGMB^{P206N}: 3 mg/mL, NEO1_{FN56}: 2 mg/mL and for the complexes, the proteins were mixed in a 1:1 ratio with a final concentration of 6 mg/mL. For the pH-dependent experiments, runs were conducted at pH 7 and pH 4. The samples were held in 3 mm path length Epon sector-shaped 2-channel centrepieces and were spun at 40,000 rpm, with 50 sample distribution scans being taken increments of 6 minutes apart. Data were collected using 280 nm absorbance optics. Data were analysed using the Sedfit package[279], which is available from <http://www.analyticalultracentrifugation.com>. The scans 6-50 were used in the continuous $c(s)$ distribution analysis, they were performed with a floating frictional ratio and baseline, $s_{\min} = 0.5$, $s_{\max} = 10$, and a resolution of 100. A value of 0.73 ml/g was used for the partial specific volumes. Graphs were plotted and Gaussian distributions fitted to the curves using the plotting program PROFIT (Uetikon am See, Switzerland).

2.13 Neurite growth assays

Cerebellar external granule layer (EGL) explants were grown on RGMB coated coverslips. First, coverslips were coated with poly-D-lysine (100 μ g/ml), washed and air-dried. Purified RGMB^{WT}, RGMB^{P206N}, RGMB^{A186R} or Fc control protein was then added to the coverslip at 50 μ g/ml mixed with laminin (40 μ g/ml, Invitrogen) in Neurobasal medium (NB, Gibco) overnight at 4 °C. Shortly before plating the explants, coverslips were washed once in NB and incubated at 37 °C with culture medium (NB with penicillin/streptomycin, l-glutamine, 18 mM HEPES and 1xB27). Explants were obtained from postnatal day (P)9 mouse pups. In brief, the cerebellum was removed and dissected in ice cold L15, and

coronal slices were cut using a MCillwain tissue chopper. Slices were further dissected to isolate the EGL and equally sized tissue explants were cut and placed onto the coated coverslips. After 3 days *in vitro* (DIV), the explants were fixed with 4 % PFA for 20 min at room temperature. For immunohistochemistry, coverslips were washed with PBS and incubated in blocking buffer (PBS with 5 % normal goat serum, 1 % BSA, 1 % glycine, and 0.4 % Triton-X100) for 1 hour at RT. Primary (mouse anti β III-tubulin, T8660 Sigma) and secondary (goat anti-mouse Alexa-488, A11029 Molecular Probes) antibodies were diluted in blocking buffer and incubated overnight at 4 °C. Nuclei and F-actin were stained using DAPI (Invitrogen) and Phalloidin-TRITC (Sigma-Aldrich), respectively, diluted in PBS. Pictures were taken with a Zeiss Scope A1 microscope with a 10x objective (Zeiss) and an Axiocam Mrm camera (Zeiss). Neurite outgrowth from the explants was analysed using ImageJ. The distance between explant and the growth cones of >20 neurites was measured (three separate experiments, total explants analysed per condition: RGMB^{WT} n=27, RGMB^{P206N} n=24, RGMB^{A186R} n=26, control n=23) and neurite lengths were normalised to the average neurite length of control explants per experiment. Three bins (short, medium, long) containing an equal proportion of measurements ranging from shortest to longest were created to establish the distribution of neurite lengths per explant. The percentage of neurites in each bin was calculated for each explant to obtain the average distribution of neurite lengths per condition. A two-way ANOVA with Bonferroni post-test was used to compare the distributions of neurite length between the experimental conditions. Experiments performed by Susan van Erp in the laboratory of R. Jeroen Pasterkamp.

Chapter 3

The Repulsive Guidance

Molecule (RGM)-Neogenin
signalling hub.

3.1 Abstract

RGMs signal through their receptor NEO1 to mediate a wide range of effects from axon guidance to leukocyte migration and apoptosis; however the molecular mechanism of RGM-NEO1 signalling has not been characterised. Recent crystal structures of RGMb in complex with NEO1 provided the first structural insight into this signalling interaction. In this chapter I present a structural analysis of these crystal structures that identified a 2:2 stoichiometry form of the RGM-NEO1 complex and allowed the design of mutations to specifically disrupt this interface. I outline AUC experiments that demonstrated the existence of the 2:2 complex in solution and validated the observed crystallographic interface. Furthermore, I show data demonstrating that this 2:2 complex is essential for full activation of RGM signalling through NEO1 in neurite growth inhibition assays. In addition, it appears that the formation of this complex is pH-dependent and that NEO1 receptors pre-cluster in the membrane. Taken together, these data allow a model for RGM signalling through NEO1 to be proposed, in which an RGM-stapled NEO1 complex acts as a pH-sensitive hub at the intersection of multiple signalling pathways.

3.2 Introduction

3.2.1 RGM binds NEO1 at the FN5-6 domains

RGMs are known to signal through their transmembrane receptor NEO1 to mediate a wide range of functions, the best characterised being the induction of axon growth cone collapse. However, the architecture of the active RGM-NEO1 signalling complex and the mode of signal transduction across the membrane are not known. Furthermore, there was no structural information available for

the RGM family, other than a predication of a partial vWFD domain in the centre of the protein. Original work, that identified NEO1 as the RGM receptor, showed that the FN1-6 domains of NEO1 were sufficient for high affinity RGM-binding[141]. More recently, after subjecting the NEO1 ectodomain to proteolytic cleavage, the region of NEO1 required for high affinity binding to human RGMC was further defined as the FN5-6 domains[280]. SPR experiments from this study revealed that RGMC binds to NEO1 FN5-6 domains with a K_d of 120 pM, with the FN6 domain making the largest contribution to this affinity; the K_d for the interaction between RGMC and FN6 alone was 20 nM[280]. They were also able to demonstrate using MALS that NEO1-FN56 and RGMC exist as monomers in solution and bind in a 1:1 stoichiometry, thereby shedding some light on the nature of RGMC-NEO1 binding.

The same group also published a high-resolution crystal structure of the FN56 domains of NEO1[148]. In the structure the two FN domains adopt the canonical FN(III) fold, comprising two antiparallel β -sheets, and adopt an extended conformation that is stabilised by a network of hydrogen bonds between the FN5 and FN6 domains. It was also noted that the NEO1 FN5 domain is highly positively charged and the authors propose a model whereby this domain interacts with either the phospholipid head groups of the membrane or a negatively charged polysaccharide chain, such as heparan sulphate, in the extracellular matrix, and thereby presents the FN6 domain for RGM binding.

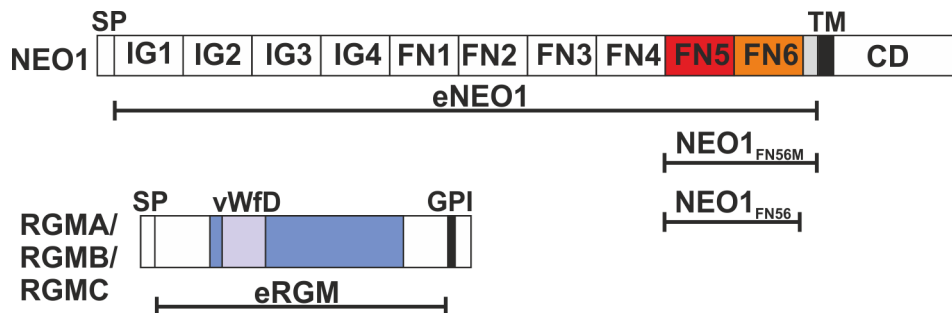


Figure 3.1: RGMB and NEO1 construct design. From [240], Science, reprinted with permission from AAAS. Schematics showing the domain organisation of human NEO1 and human RGMB. Black bars indicate the crystallisation constructs eRGMB and NEO1_{FN56}. The portions of the protein resolved in the crystal structure are coloured: the RGMB middle domain (purple), including the partial von Willibrand Factor D domain (vWFD) (lilac), and the NEO1 FN5 (red) and FN6 (orange) domains. SP = signal peptide, IG = immunoglobulin-like domain, FN = fibronectin III-like domain, TM = transmembrane helix, CD = cytoplasmic domain, GPI = glycosylphosphatidylinositol anchor. (Figure taken from [240]).

3.2.2 The structure of the RGM-NEO1 complex

This prior structural and binding data was used in our laboratory to guide the expression of constructs of soluble full-length RGMB (eRGMB) (i.e. without the GPI-anchor) and the FN5-6 domains of NEO1 (NEO1_{FN56}) (Fig. 3.1). These proteins were mixed in a 1:1 ratio for crystallisation trials and a series of eRGMB-NEO1_{FN56} complex crystal structures were solved [240]. In these structures the middle domain of RGMB interacts with both the NEO1 FN5 and FN6 domains (Fig. 3.2a). The dominant contact is between RGMB and one face of the antiparallel β -sheet of the NEO1 FN6 domain, which contains β -strands A, B and E. However, the L3 loop of the NEO1 FN5 domain, located between β -strands C and C', also contributes to the interaction. In total the interface is formed of 17 hydrogen bonds and 147 non-bonded contacts (Fig. 3.2b). SPR experiments verified that the NEO1 FN56 domains are sufficient for the NEO1-RGMB interaction, as described in earlier studies [141] (Fig. 3.2c,d). Mutagenesis of several key residues at the interface, followed by SPR binding analysis, validated the

observed complex structure (Fig. 3.2b), and one representative example of this is shown: the NEO1 mutation L1046E totally abolished RGMB-NEO1 binding (Fig. 3.2e). A further validation of the observed RGMB-NEO1 interaction is that the RGMC mutation G320V is located close to the interface. This mutation is linked to the blood iron overload disorder JHH and is known to disrupt RGMC binding to NEO1[190]. In contrast, mutations in the L3 loop of the NEO1 FN5 domain reduced but did not abolish binding[240], suggesting that this loop is important but not essential for the interaction. Interestingly, the L3 loop is very poorly conserved between NEO1 and its close homologue DCC, unlike the majority of the interface; and given that the binding of RGMs to DCC is approximately one thousandth of that for NEO1[240], it may therefore act as a specificity determinant for the RGM-NEO1 interaction.

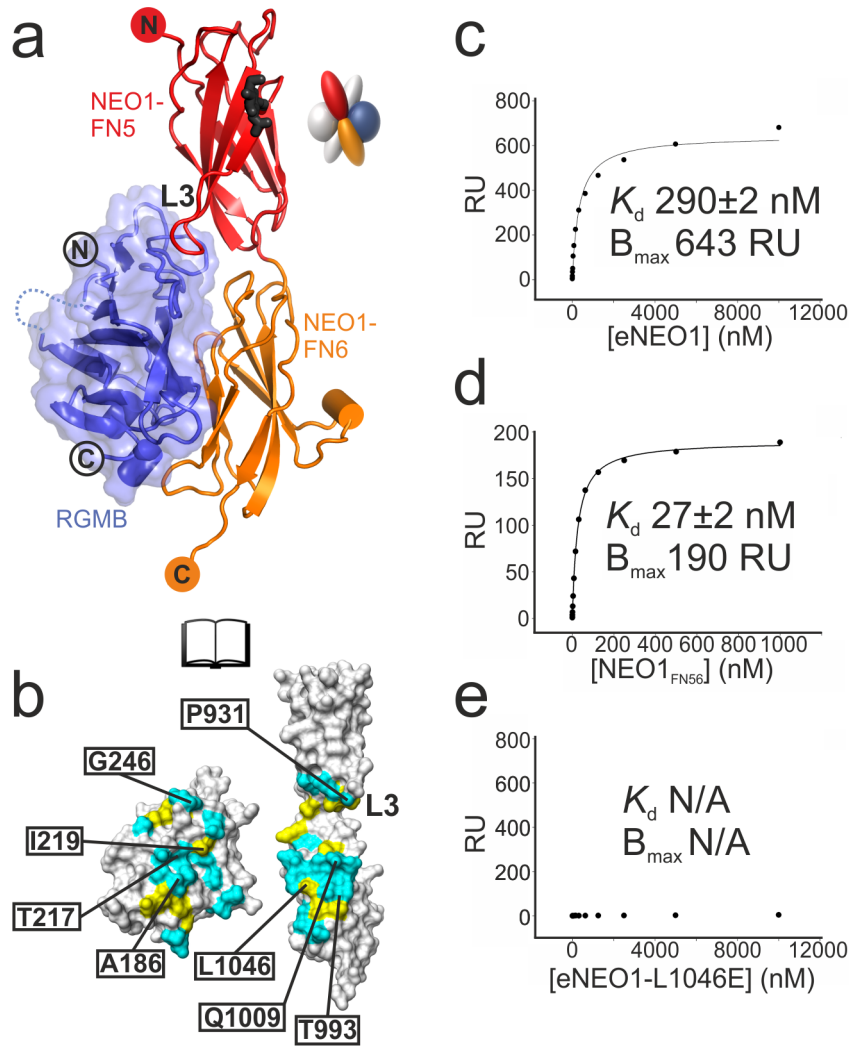


Figure 3.2: Detailed interactions of the RGM-NEO1 complex. From [240], Science, reprinted with permission from AAAS. (a) Ribbon representation of eRGM-NEO1_{FN56} complex. RGM transparent surface is shown, colour-coding is as Fig. 3.1 and N- and C-termini are marked. The L3-loop of NEO1 is labelled. (b) Open-book view showing the solvent accessible surface of the eRGM-NEO1_{FN56} interface (formed by 17 hydrogen bonds and 147 non-bonded contacts). Cyan = hydrophilic interactions; yellow = non-bonded contacts. Residues analysed for their affect on eRGM-NEO1_{FN56} binding, by site-directed mutagenesis followed by SPR, are labelled[240] (c-e) SPR equilibrium-binding analysis. Different concentrations of the full-length NEO1 ectodomain (eNEO1) (c), NEO1_{FN56} (d) and eNEO1-L1046E (e) were injected over surfaces coupled with eRGM. The K_d and B_{max} of the curve fit are shown. (Figure taken from[240]).

3.2.3 The structure of the RGM middle domain

In all of these eRGMB-NEO1_{FN56} complex structures only the middle domain of RGMB could be resolved, comprising residues 134-338, thereby identifying this as the region of RGM required for NEO1 binding. This region adopts a novel fold consisting of a tightly packed β -sandwich extended by four short helices at the C-terminus (Fig. 3.3). The structure is stabilised by three disulphide bonds that appear to orient the N- and C- termini away from the middle domain and into solvent channels in the crystal. This suggests that the N- and C-terminal domains of RGMB, which were included in the crystallisation construct, are not associated with the middle domain (Fig. 3.3b). The conserved RGM autocatalytic cleavage site between Asp168-Pro169 is located in the loop connecting β -sheets 1 and 2 (Fig. 3.3 and 3.4a).

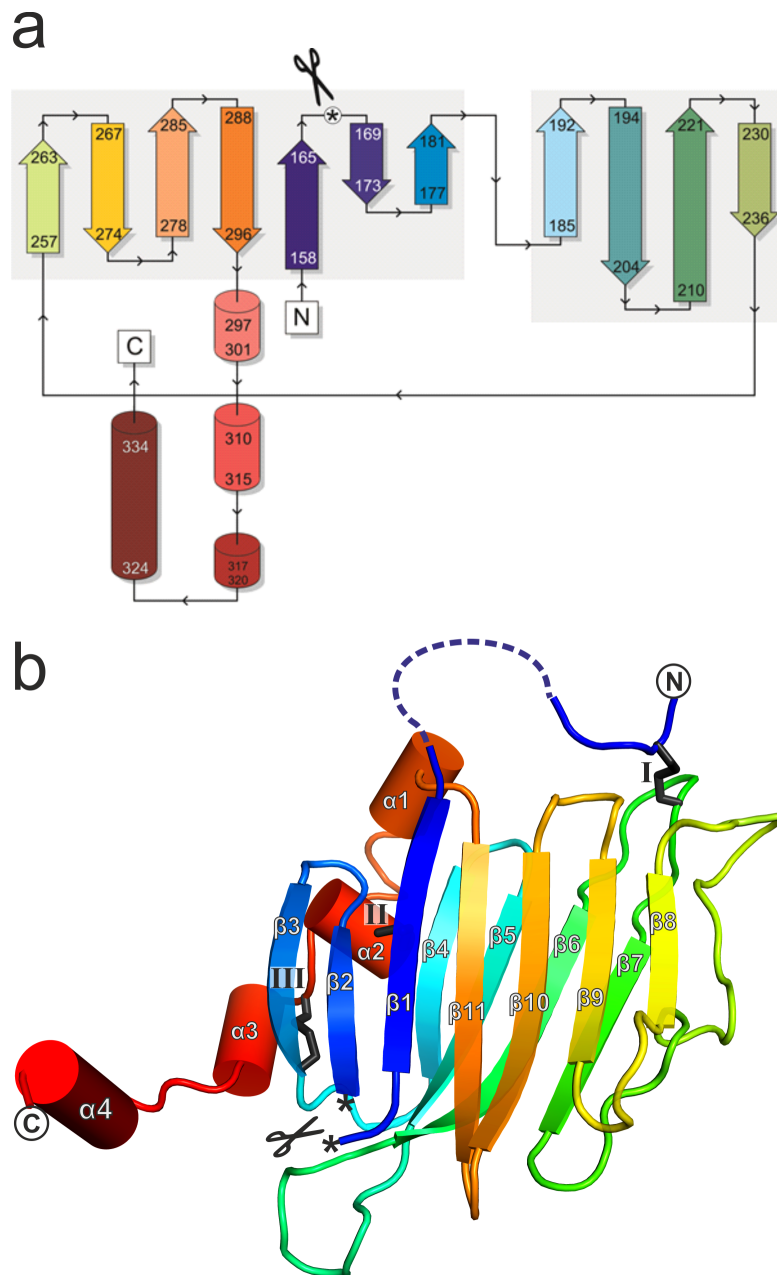


Figure 3.3: The structure of the RGM middle domain. From [240], [Science](#), reprinted with permission from AAAS. **(a)** Topology diagram of the middle domain of RGM in rainbow colouring (blue = N-terminus, red = C-terminus). The autocatalytic cleavage site is highlighted by an asterisk and cartoon scissors. The figure is adapted from PDBSUM (<http://www.ebi.ac.uk/pdbsum/>). Residue numbering corresponds to full-length human RGM (including the signal peptide). Grey boxes highlight the β -sheets. **(b)** eRGM ribbon diagram, colouring as in (a). Disulphides (black sticks) are depicted with Roman numerals. The autocatalytic cleavage site is marked with an asterisk and cartoon scissors. A dotted line marks the disordered loop connecting the N-terminus with the $\beta 1$ -strand. RGMs comprise a novel protein fold consisting of a tightly packed β -sandwich with seven strands on one sheet and four strands on the other, extended by four short helices at the C-terminus. (Figure taken from [240]).

These acid-labile Asp-Pro bonds are known to be hydrolysed in low-pH environments within the cell, such as in the Golgi or secretory vesicles[281]. For RGMs, this has been shown to occur by autocatalytic cleavage[42]. Mechanistically this involves intramolecular catalysis by aspartate carboxylate anion displacement of the protonated nitrogen of the peptide bond. Since the prolyl nitrogen has the strongest basicity of all amino acids, this enhances the peptide bond cleavage by increasing protonation of the leaving group, which is favoured in lower pH environments. Due to this cleavage reaction RGMs consist of two protein chains held together by disulphide bonds, and this is evident when the protein is analysed by reducing and non-reducing SDS-PAGE (Fig. 3.4b). The autocatalytic cleavage allows for Pro169 to be deeply buried in the hydrophobic core of the protein, which contributes to the stability of the overall fold (Fig. 3.4a). The residues of the autocatalytic cleavage site are conserved across various vWFD family members and similar autocatalytic cleavage has been functionally characterised in Mucins[282, 58] and BMPER[283] (Fig. 3.4c). However, RGM family members are missing the Gly-Leu-Cys-Gly motif that is normally located towards the C-terminus of the canonical vWFD domain[145], which makes it difficult to predict whether the other vWfD family members will adopt a similar fold to the RGMs.

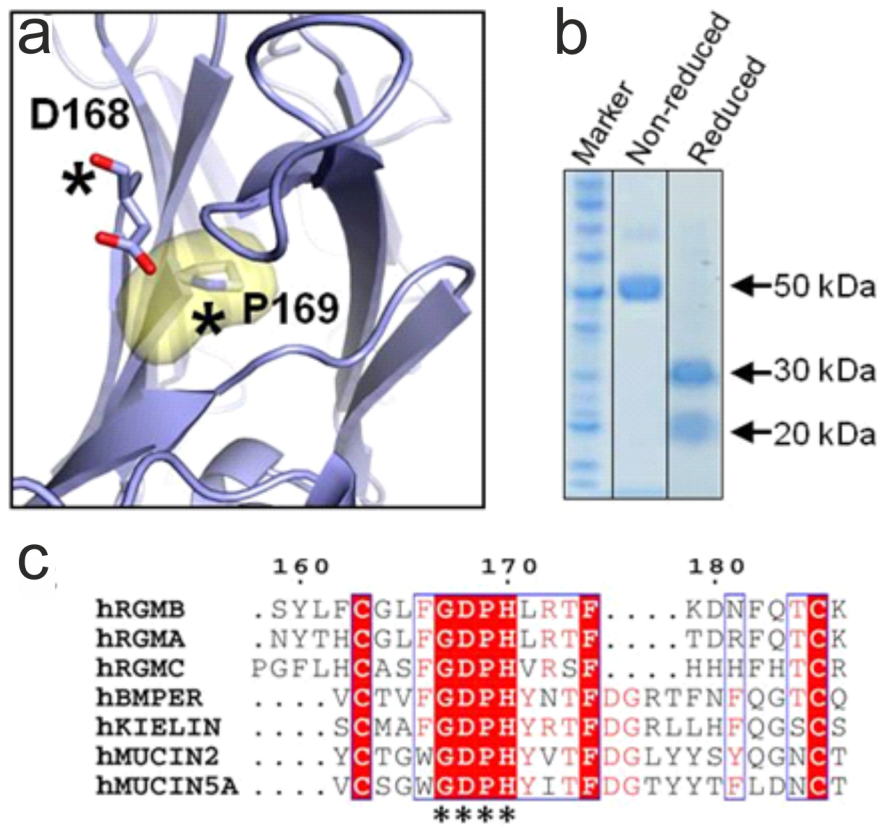


Figure 3.4: The RGM autocatalytic cleavage site. From [240], [Science](#), reprinted with permission from AAAS. **(a)** Close-up view of the RGM autocatalytic cleavage site. The amino acids involved are labelled. The yellow sphere shows the solvent accessible surface of Pro169. The asterisk indicates the site of cleavage. **(b)** Reducing and non-reducing SDS-PAGE of SEC-purified human RGMA. All transiently expressed and secreted RGMA and RGM proteins from human HEK293T cells are ~100 % processed. **(c)** Sequence alignment of vWFD family members (Human BMPER: Uniprot Q8N8U9, human Kielin: UniProt Q6ZWJ8, human Mucin 2: UniProt Q02817, human Mucin 5A: UniProt P98088) showing conservation of the RGM autocatalytic cleavage site ("GDPH"). Sequences were aligned using MULTALIN (bioinfo.genotoul.fr/multalin/multalin.html) and formatted with ESPRIPT (espript.ibcp.fr/ESPrIPT/ESPrIPT/). Numbering corresponds to the full length human RGM (including the secretion signal). Asterisks highlight the conserved autocatalytic cleavage site residues. (Figure taken from [240]).

3.2.4 A higher order RGMB-NEO1 complex?

The structure of the eRGMB-NEO1_{FN56} complex was solved in three independent crystal forms, two at neutral pH and one at low pH[240] (Fig. 3.5). Remarkably, in both of the neutral pH forms the asymmetric unit was composed of two molecules of NEO1_{FN56} and two molecules of eRGMB, with the arrangement of this 2:2 complex preserved between crystal forms. Furthermore, the architecture of this complex appears to place the NEO1 C-termini in close proximity, hinting at a potential mechanism of RGM-induced dimerisation of NEO1, which could be important for signalling. In contrast, the crystal form at low pH showed only a 1:1 complex.

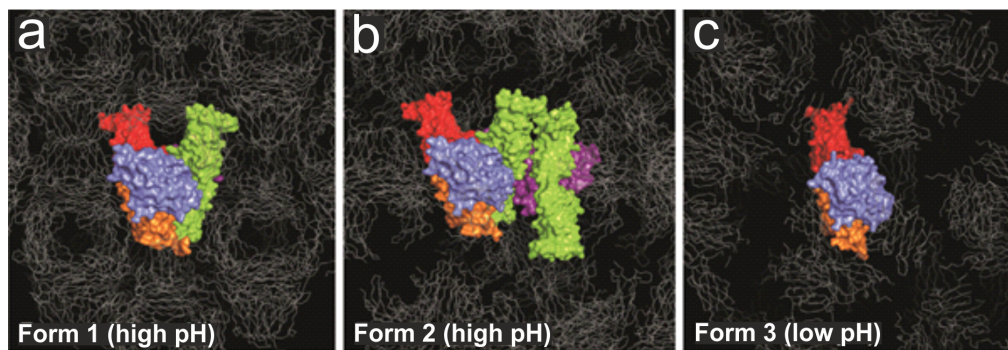


Figure 3.5: Crystal packing of the different eRGMB-NEO1_{FN56} complex structures. From [240], [Science](#), reprinted with permission from AAAS. (a) eRGMB-NEO1_{FN56} Form 1, (b) eRGMB-NEO1_{FN56} Form 2 and (c) eRGMB-NEO1_{FN56} Form 3. RGMB: purple/violet; NEO1: red (FN5)/orange (FN6) and green. The neutral pH crystal forms 1 and 2 contain a 2:2 eRGMB-NEO1_{FN56} complex in the asymmetric unit. The low pH crystal form 3 contains only a 1:1 complex in the asymmetric unit. (Figure taken from[240]).

MALS experiments using the purified eRGMB and NEO1_{FN56} proteins revealed that the proteins are monomeric and only a 1:1 stoichiometry complex could be observed (Fig. 3.6); which confirmed observations from another study[280]. Nevertheless, it is commonplace in biology for active signalling complexes to be built from low affinity, transient interactions to maintain dynamic and responsive systems. It is, therefore, perhaps expected that this experiment did not reveal

the 2:2 complex form observed in the crystal structures.

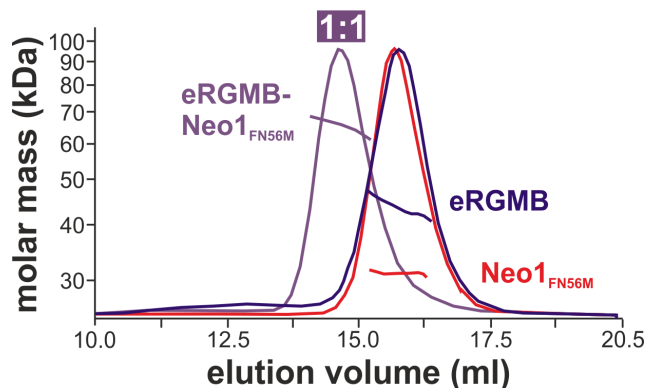


Figure 3.6: Preliminary analysis of the oligomeric state of the eRGMB-NEO1_{FN56} complex. From [240], [Science](#), reprinted with permission from AAAS. Multi-angle light scattering (MALS) analysis of the eRGMB-NEO1_{FN56} complex. eRGMB (blue) and NEO1_{FN56} (red) behave as monomers with the following molecular weights (measured/theoretical): eRGMB: $42.64 \pm 0.1 / 43.35$ kDa; NEO1_{FN56}: $31.04 \pm 0.1 / 31.05$ kDa. The eRGMB-NEO1_{FN56} complex (magenta) behaves as a complex of one RGMB and one NEO1 (molecular weight measured/theoretical: $69.04 \pm 0.1 / 74.40$ kDa). Theoretical molecular weights of glycosylated proteins were calculated using PROTPARAM at the EXPASY server (www.expasy.ch). (Figure taken from [240]).

3.2.5 Chapter outline

Building on these preliminary studies of the RGMB-NEO1 interaction, the aim of this chapter was to use the crystal structures as a basis for probing the molecular mechanism of RGM signalling through NEO1. Towards this an analysis of the eRGMB-NEO1_{FN56} complex structures was performed. The locations of JHH-linked mutations in RGMC were mapped onto the corresponding residues in the RGMB structure and these mutants were tested for secretion in a mammalian cell protein expression system. Most of these mutations inhibited the secretion of RGMB, providing a molecular rationale for the disease-causing mechanism of these mutations. An analysis of the interactions observed in the eRGMB-NEO1_{FN56} 2:2 complex crystal form identified two key interfaces, termed “site-1” and “site-2”. Optimisation of AUC experiments allowed sampling of a much higher concentration range of the two proteins than had been used in the prelimi-

nary MALS experiments. These revealed that an eRGMB-NEO1_{FN56} 2:2 complex exists in solution, and in fact that this complex form is pH sensitive. Mutants at the site-2 interface were designed for AUC experiments to validate the complex structure. Furthermore, eRGMB site-1 and site-2 mutants were used in neurite outgrowth inhibition assays to assess the functional relevance of the 2:2 complex form for RGM-activated NEO1 signalling. These experiments showed that the 2:2 complex is important for fully active NEO1 signalling and this, combined with data suggesting that NEO1 pre-clusters in the cell membrane, allowed us to propose a mechanism for RGM signalling through NEO1 that incorporates dimerisation and an endocytosis-linked mechanism of signal termination.

3.3 Results

3.3.1 JHH-linked mutations abolish the secretion of RGMC

Over 14 different mutations in RGMC can lead to the disease JHH, an early onset blood iron overload disorder. A sequence alignment of vertebrate RGM family members was used to infer the residues in RGMB corresponding to these mutation sites, many of which were highly conserved (Fig. 1.6). Ten JHH-linked mutations in total map onto the β -sandwich region in the middle domain of RGMB, with the remainder localised to the N-terminal domain not visible in the eRGMB-NEO1_{FN56} structures (Fig. 3.7a). Strikingly, seven of these cluster around the autocatalytic cleavage site, including the RGMC mutation D172E, which affects the catalytic aspartic acid of the cleavage site itself (Fig. 3.7a). All of the JHH-linked RGMC mutants were tested for secretion in a mammalian cell expression system using small-scale expression tests. Using SDS-PAGE, the protein distribution in the secreted and cellular samples, under both reducing and non-reducing conditions, was analysed by Western blotting with an anti-His₆

antibody (Fig. 3.7b,c). As expected, under reducing conditions the RGMC migrated as a double band corresponding to the processed and unprocessed forms. Interestingly, 12 out of 14 of the mutations totally abolished secretion of RGMC, although total protein levels seemed to be unaffected (Fig. 3.7b,c). This highlights the importance of the cleavage site for the structural integrity of the β -sandwich and shows that the stability of this domain is crucial for viable secretion of the entire protein. Only two mutations were secreted at similar levels to wild-type RGMC, G99R and L101P, and these are located in the N-terminal portion of RGMC and do not appear to affect the stability of the protein (Fig. 3.7b,c).

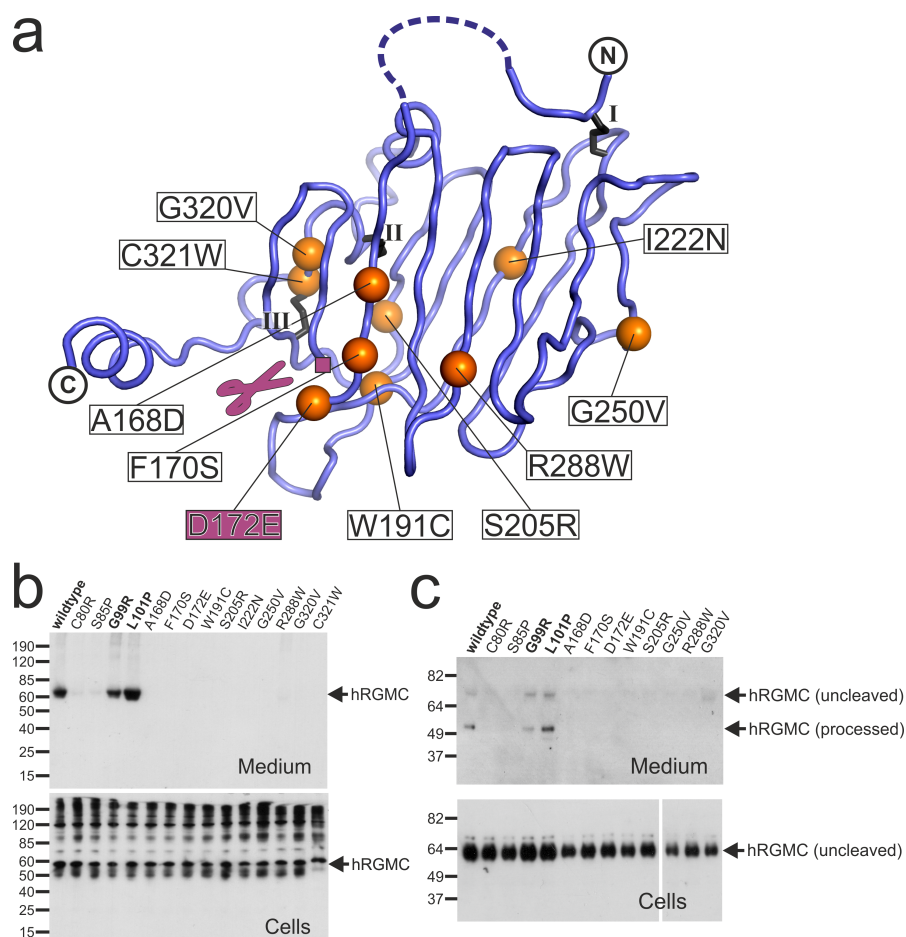


Figure 3.7: Structural and functional characterisation of RGM C mutations causing JHH. From [240], [Science](#), reprinted with permission from AAAS. (a) RGM C JHH-linked mutations mapped onto the RGM C structure. Equivalent residues in RGM C are shown as orange spheres. The autocatalytic cleavage site is highlighted and the D172E mutation is outlined by a pink box. N- and C-termini are labelled; disulphide bonds are labelled with Roman numerals; the disordered loop is shown as a dotted line. (b-c) Western blots of small-scale expression tests for the disease-related RGM C mutants. Top panels: media samples; lower panels: cell samples. Both non-reducing (b) and reducing (c) SDS-PAGE analysis was performed.

3.3.2 A structural analysis of the 2:2 eRGMB-NEO1_{FN56} complex

An analysis of the eRGMB-NEO1_{FN56} 2:2 stoichiometry complex revealed that the two molecules of NEO1 contact each other with a buried surface area of 835 Å², whereas the two RGMB molecules in the complex do not touch (Fig. 3.8). Two eRGMB-NEO1_{FN56} interfaces are clearly identifiable. One larger interface, termed the “site-1” interface, with a buried surface area of 2350 Å², is preserved in the 1:1 complex; and a smaller interface termed the “site-2” interface with a buried surface area of 695 Å², is responsible for mediating 2:2 complex formation (Fig. 3.8).

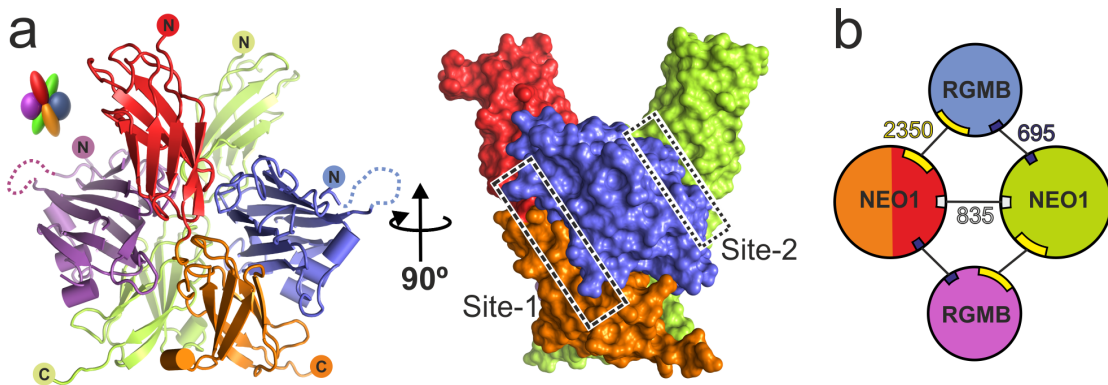


Figure 3.8: The structure of the 2:2 eRGMB-NEO1_{FN56} complex. From [240], *Science*, reprinted with permission from AAAS. (a) Ribbon (left panel) and surface representation of the 2:2 eRGMB-NEO1_{FN56} complex. Site-1 and site-2 interfaces are highlighted with boxes. eRGMB: blue/violet; NEO1_{FN56}: red (FN5)/orange (FN6) and green. Right panel is 90° rotated around the y-axis compared to the left panel representation. (b) Schematics of the 2:2 eRGMB-NEO1_{FN56} complex showing the protein-protein contacts. Interface buried surface areas (Å²) are shown, calculated using the PISA webserver[254]. Colour-coding is as (a). (Figure from [240]).

The site-2 interface, similar to the site-1 interface, is composed of interactions between the middle domain of RGMB and both the FN5 and FN6 domains of NEO1. RGMB utilises two loop regions between the β -strands β 5- β 6 and β 10- β 11, which make contact with the N-terminal portion of the NEO1 FN6

domain and parts of the L3 loop of the FN5 domain (Fig. 3.9a). The site-2 interface comprises 1 hydrogen bond and 25 non-bonded contacts and is largely hydrophobic in nature; for example, Pro206, which is located at the centre of the site-2 interface and forms a hydrophobic interaction (Fig. 3.9a,b). The residues making up the interface are well conserved across NEO1 and RGM vertebrate orthologues and paralogues suggesting that this interface is unlikely to be a crystal-packing artefact (Fig. 3.9c).

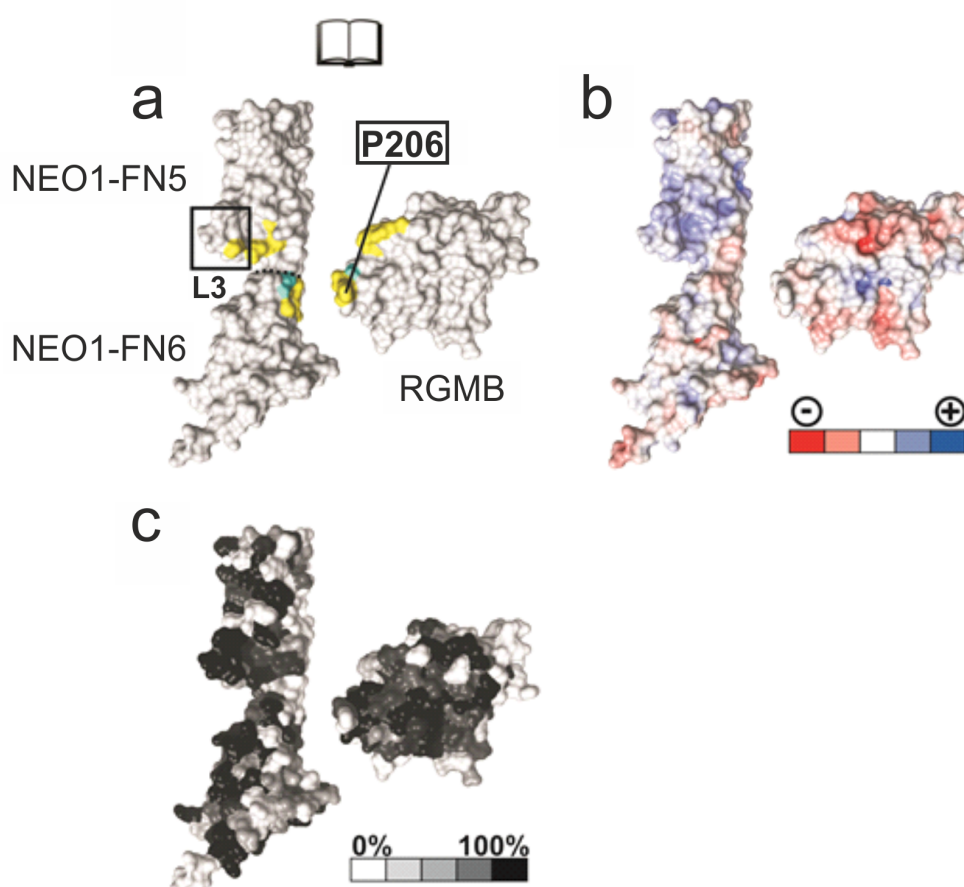


Figure 3.9: The eRGM-NEO1_{FN56} site-2 interactions. From [240], *Science*, reprinted with permission from AAAS. The eRGM-NEO1_{FN56} complex is shown in surface representation in an open book view. (a) Site-2 interface residues are highlighted (non-bonded interactions: yellow; hydrophilic interactions: cyan). The NEO1 FN5 domain L3 loop is marked. RGM residue P206, located directly at the interface is labelled. (b) Residues are coloured according to electrostatic potential, from red (-8 kBT/ec) to blue (+8 kBT/ec), showing that the interface is largely hydrophobic. (c) The complex is colour-coded according to residue conservation (from non-conserved, white, to conserved, black) based on alignments containing sequences from 17 NEO1 orthologues and 94 RGM orthologues and paralogues from vertebrates. (Figure from [240]).

3.3.3 AUC analysis of eRGMB and NEO1_{FN56}

Previous studies have analysed the stoichiometry of the RGM-NEO1_{FN56} complex using MALS and found it to be a 1:1 interaction, with both proteins appearing monomeric in isolation[240, 280]. However, these experiments analysed the complex formation at relatively low protein concentrations, for example, the analysis from our laboratory was conducted at a protein concentration of 3 μ M. This motivated the use of AUC to analyse the stoichiometry of the eRGMB-NEO1_{FN56} complex, as this technique allows sampling of much higher protein concentrations that mimic the crowded local protein distribution found in the restricted two-dimensional diffusion environment of the plasma membrane. Towards this, eRGMB and NEO1_{FN56} were expressed and the protein isolated by metal ion affinity and purified using SEC (Fig. 3.10a,b). Both proteins were soluble and monodisperse; eRGMB exhibited a single peak in the SEC trace and analysis by non-reducing SDS-PAGE showed the protein to be largely pure with very minor degradation products (Fig. 3.10a). The SEC trace for NEO1_{FN56} showed a single peak with a shoulder towards lower molecular weights and an SDS-PAGE analysis revealed multiple bands, most likely due to different glycosylation species (Fig. 3.10b). Therefore to increase the homogeneity of the NEO1_{FN56} sample, only SEC fractions 3-6 were pooled for subsequent analysis (Black bar, Fig. 3.10b). The individual protein components were concentrated, eRGMB to 3 mg/ml and NEO1_{FN56} to 2 mg/ml, and an SV-AUC experiment was performed to determine whether the protein complexes were likely to be suitable for analysis by AUC (Fig. 3.10c). The distribution plots, obtained from the fitting of SV data using a continuous $c(s)$ distribution model, exhibited a Gaussian distribution suggesting that both proteins exist in solution as discrete monomers (Fig. 3.10c). This confirmed that the eRGMB and NEO1_{FN56} constructs were well suited to SV-AUC analysis and allowed investigation into the formation of the complex to proceed.

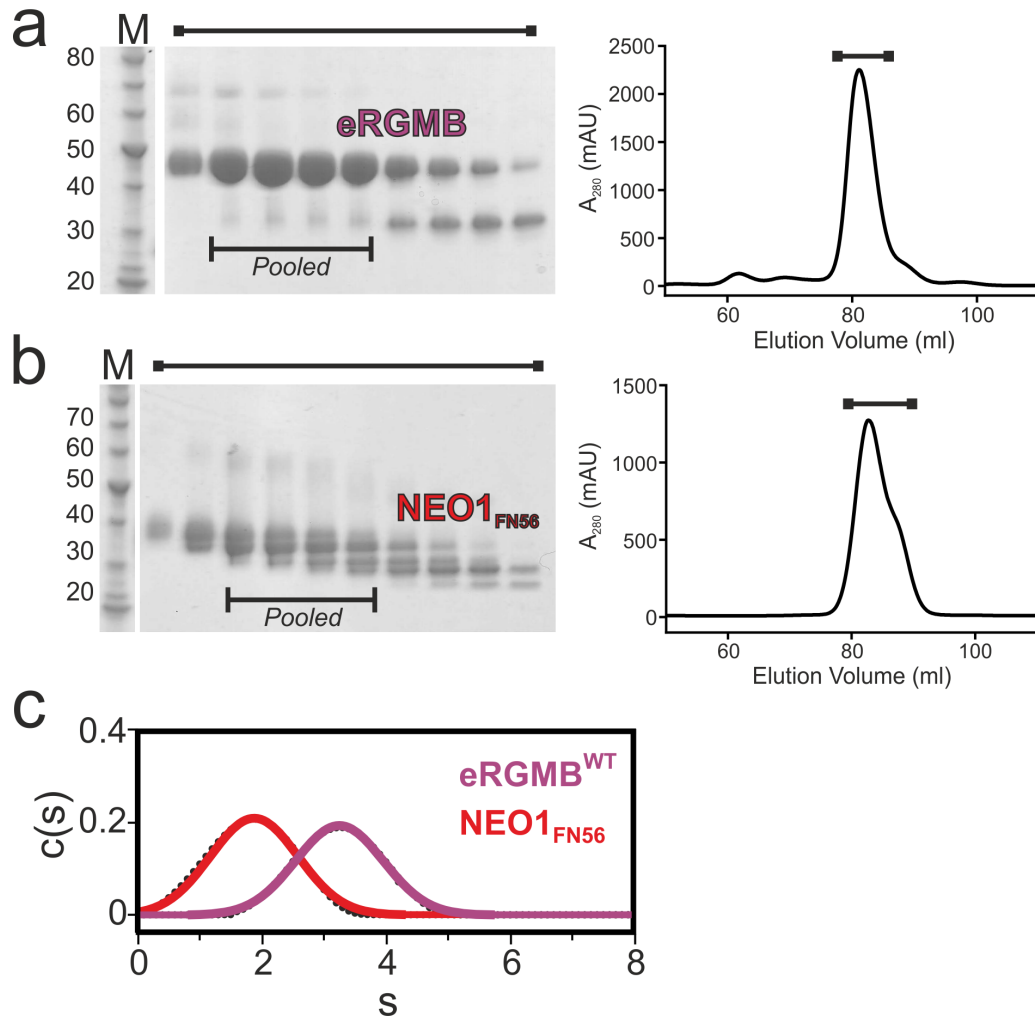


Figure 3.10: Purification and AUC analysis of eRGMB and NEO1_{FN56}. (a-b) Left panel shows Coomassie Blue stained, non-reducing (a) and reducing (b) SDS-PAGE gels of fractions from SEC-purified His₆-tagged eRGMB (a) and NEO1_{FN56} (b). M = marker, molecular weights = kDa. The right panels show the corresponding SEC profiles. A black bar indicates SEC fractions analysed by SDS-PAGE. Fractions pooled for further experiments are indicated. (c) SV-AUC analysis of eRGMB (purple) and NEO1_{FN56} (red). Distribution plots obtained from the fitting of SV data using a continuous $c(s)$ distribution model (solid line Gaussian fit to data plotted as symbols) are shown.

3.3.4 Design, expression and purification of a site-2 mutant

Towards this AUC analysis of the eRGMB-NEO1_{FN56} complex, mutants were designed to specifically disrupt the site-2 interface observed in the crystal structures. These would then be used to verify that any higher order oligomers observed were due to the site-2 interaction, allowing validation of the architecture of the 2:2 crystal complex. An analysis of the site-2 interface revealed P206^{RGMB} and V204^{RGMB} to be two key interacting residues located at the centre of the RGMB interface with the NEO1 FN6 domain. Residue V204 forms a hydrogen bond with T986^{NEO1} and P206 makes hydrophobic interactions with P1073^{NEO1} (Fig. 3.11a,b). Site-directed mutagenesis of eRGMB was used to clone a V204G/P206G double mutant into the pHLsec vector background for expression[200]. This mutant construct was designed to abolish these interactions and consequently disrupt the formation of the 2:2 complex through the site-2 interface.

Four other eRGMB mutations were also generated that introduced a bulky N-linked glycan close to the position of the site-2 interface to sterically disrupt the association of eRGMB and NEO1_{FN56} and prevent the formation of the 2:2 complex. The consensus sequence for mammalian N-linked glycosylation is N-X-S/T (where X can be any amino acid and Thr is more common than Ser), so mutations were targeted to introduce novel consensus glycosylation sites into RGMB. This included the mutation I216T, which introduced an N-linked glycan at N214; and the mutation P206N, which introduced a glycan at the new N206 position. Two double mutants Q284N/G286S and R287N/L289T, were also cloned, which introduced N-glycans at positions N284 and N287 respectively. These four mutants were cloned into the pHLsec vector for expression in mammalian cells[200]. Small-scale expression tests revealed that the eRGMB^{Q284N/G286S} and

eRGMB^{R287N/L289T} double mutants did not express, however, the other mutants did. Therefore large-scale expression and purification of the eRGMB^{V204G/P206G}, eRGMB^{P206N} and eRGMB^{I216T} mutants was conducted. HEK293T cells were transfected with these constructs and after four days the media was dialysed against PBS pH 7.5 and the proteins were purified by Talon metal affinity followed by SEC in 10 mM HEPES, pH 7.5, 150 mM NaCl, as described in materials and methods. A representative SEC trace (right panel) and accompanying SDS-PAGE (left panel) for one of the mutant constructs, eRGMB^{P206N}, is shown (Fig. 3.11c); the yield for all three mutants was similar.

The purified eRGMB mutant proteins were subsequently concentrated to approximately 3 mg/ml for SV-AUC analysis as described in materials and methods (Fig. 3.11d-f). Both eRGMB^{P206N} and eRGMB^{V204G/P206G} were monomeric, exhibiting SV distribution curves that could be fitted with a single Gaussian (Fig. 3.11d,f). In contrast, the curve for eRGMB^{I216T} required three separate Gaussian distributions for curve-fitting, probably indicating that this protein was not stable and that several oligomeric species exist in solution (Fig. 3.11e). This mutant was therefore excluded from a subsequent analysis of the formation of the eRGMB-NEO1_{FN56} complex.

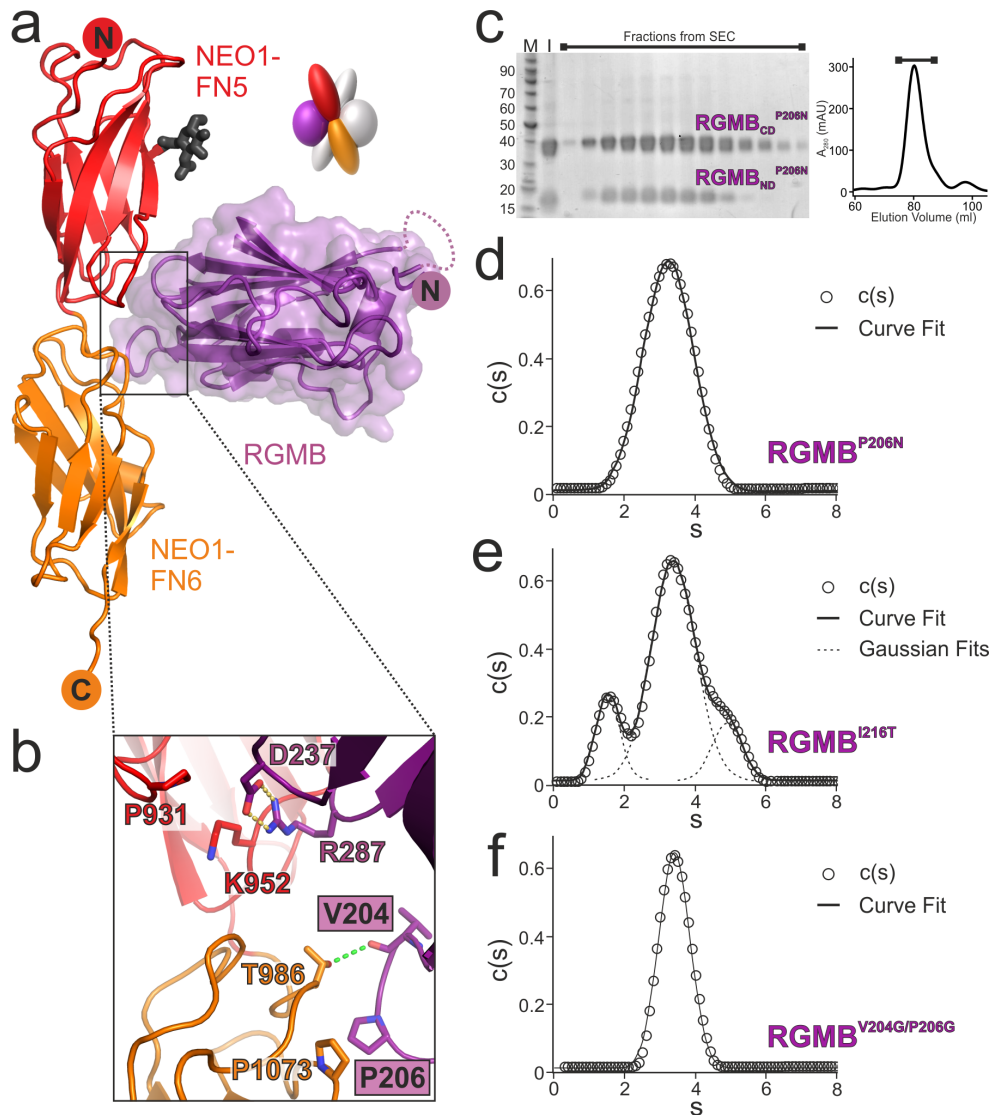


Figure 3.11: Targeting the site-2 interface. From [240], Science, reprinted with permission from AAAS. **(a)** Ribbon representation of the eRGM-NEO1_{FN56} complex showing the site-2 interface. RGM (purple) $\beta 5$ - $\beta 6$ and $\beta 10$ - $\beta 11$ loops contact the NEO1 FN5 (red) and FN6 (orange) domains. N- and C-termini are shown. A NEO1 N-linked glycan shown as black sticks. **(b)** Close-up view of the interactions at the site-2 interface, coloured as in (a), proteins are shown in ribbon representation. Key interacting residues are shown as sticks. Purple boxes highlight residues targeted for mutation. Hydrogen bonds are shown as dotted lines (yellow = intramolecular, green = intermolecular). **(c)** Reducing SDS-PAGE (left panel) and corresponding SEC profile (right panel) for RGM^{P206N}. Bands corresponding to RGM_{ND} and RGM_{CD} are labelled. Black bars indicate SEC fractions analysed by SDS-PAGE. **(d-f)** SV-AUC analysis of eRGM^{P206N} (d), eRGM^{I216T} (e), eRGM^{V204G/P206G} (f). Data points (circles) fitted using a continuous $c(s)$ distribution model (solid line). Gaussian peaks contributing to the overall distributions are shown as dotted lines.

3.3.5 The observed 2:2 eRGMB-NEO1_{FN56} complex exists in solution

To investigate the formation of the 2:2 complex using SV-AUC, eRGMB and NEO1_{FN56} were mixed in a 1:1 molar ratio and the sample concentrated to 6 mg/ml. The SV distribution for the eRGMB-NEO1_{FN56} complex could be described by fitting a curve comprising three Gaussian distributions (Fig. 3.12b). Comparison to the SV profiles for the individual components (Fig. 3.12a) allowed each Gaussian peak in the complex distribution to be assigned. The left Gaussian (shown in pink) corresponds, most likely, to free excess eRGMB, as both curves have a similar sedimentation coefficient (*s*). The middle Gaussian (shown in blue) probably corresponds to the 1:1 complex. There is a marked shift in sedimentation coefficient from that for either of the single components (from *s* ~2 and *s* ~3.2, to *s* ~4), suggesting an increase in the mass of the species consistent with a 1:1 complex. The Gaussian furthest to the right (shown in yellow) indicates a species with an increase in mass over the 1:1 complex, which most likely corresponds to the 2:2 complex (Fig. 3.12b). Although this analysis demonstrates that the population of eRGMB-NEO1_{FN56} complexes is dominated by the 1:1 stoichiometry species due to the high-affinity of this interaction, the overall curve shows a distinct shoulder, which can be attributed to the presence of a significant proportion of 2:2 oligomers.

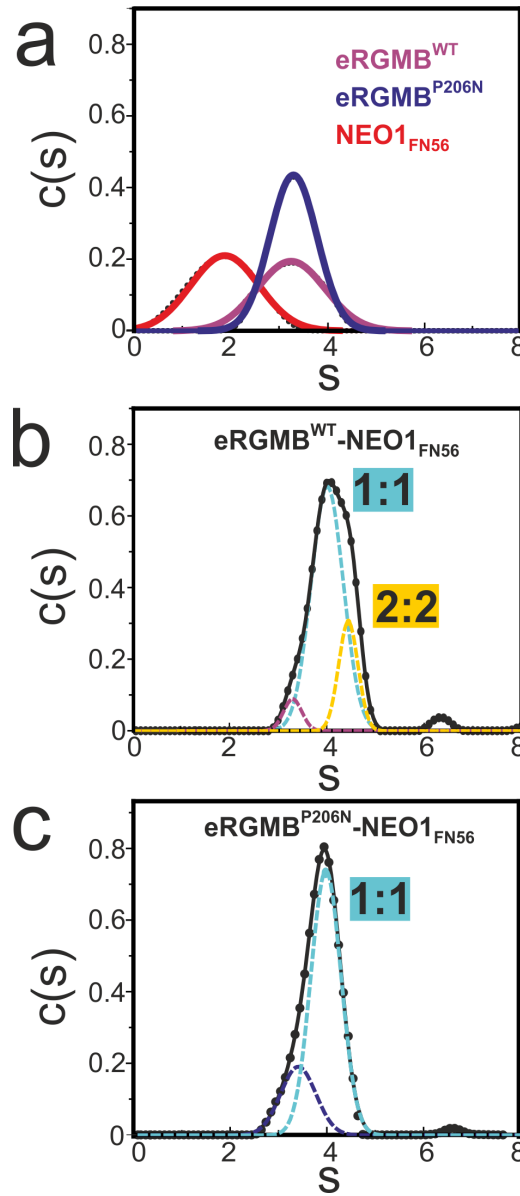


Figure 3.12: SV-AUC experiments reveal the 2:2 complex. From [240], Science, reprinted with permission from AAAS. (a) SV-AUC analysis of the single components including $eRGMB^{WT}$ (purple), $eRGMB^{P206N}$ (blue) and $NEO1_{FN56}$ (red). Individual components run as monomeric species. (b-c) SV-AUC experiments to determine the stoichiometry of the complexes, including $eRGMB^{WT}$ - $NEO1_{FN56}$ (b) and $eRGMB^{P206N}$ - $NEO1_{FN56}$ (c). Data points (black circles) fitted using a continuous $c(s)$ distribution model (black solid line). Gaussian peaks contributing to the overall distributions are displayed as dotted lines. For $eRGMB^{WT}$ - $NEO1_{FN56}$ (b, r.m.s.d. = 0.0038) these correspond to free $eRGMB^{WT}$ (purple), the 1:1 $eRGMB^{WT}$ - $NEO1_{FN56}$ complex (cyan) and the 2:2 complex (yellow). For $eRGMB^{P206N}$ - $NEO1_{FN56}$ (c, r.m.s.d. = 0.0063) these correspond to free $eRGMB^{P206N}$ (blue) and the 1:1 $eRGMB^{P206N}$ - $NEO1_{FN56}$ complex (cyan). Gaussian distributions corresponding to the 1:1 (cyan) and 2:2 (yellow) complexes are highlighted with coloured boxes.

To confirm our prediction that the Gaussian fit shown in yellow represents the 2:2 complex, the SV profiles of the eRGMB^{P206N}- and eRGMB^{V204G/P206G}-NEO1_{FN56} complexes were analysed. Whilst there was no difference between the profile for eRGMB-NEO1_{FN56} and eRGMB^{V204G/P206G}-NEO1_{FN56} (data not shown), there was a marked difference in the eRGMB^{P206N}-NEO1_{FN56} curve (Fig. 3.12c). The higher-order oligomer, indicated by the presence of a shoulder in the $c(s)$ profile, that was visible for the wild type eRGMB-NEO1_{FN56} complex was not present. The distribution curve could be accurately described by two Gaussians fits corresponding to free eRGMB^{P206N} and the 1:1 eRGMB^{P206N}-NEO1_{FN56} complex. This analysis suggests that the double glycine mutation V204G/P206G was not sufficient to disrupt the eRGMB-NEO1_{FN56} interaction. However, the introduction of a glycan at position N206 in the centre of the site-2 interface was sufficient to disrupt the formation of the 2:2 complex, but did not affect the assembly of the 1:1 complex. These experiments support our assignment of the 2:2 complex peak in our AUC analysis and validate the site-2 interface from the crystal structure.

3.3.6 Formation of the 2:2 complex is pH-dependent

Of the three crystal forms of the eRGMB-NEO1_{FN56} complex solved, the 2:2 arrangement was only observed at neutral pH; the crystal structure solved at pH 4.5 contained only the 1:1 complex in the asymmetric unit (Fig. 3.5). Structural analysis of the site-2 interface identified several residues that are charged, including D237^{RGMB}, R287^{RGMB} and D899^{NEO1} and numerous interactions, coordinating water molecules and polar amino acid side chains (e.g. H898^{NEO1}) that could be sensitive to pH changes (Fig. 3.13a,b). This analysis suggested that the formation of the 2:2 complex might be pH-dependent.

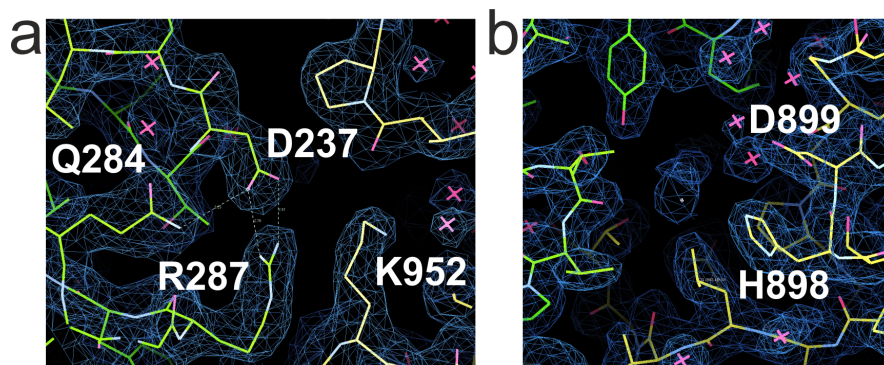


Figure 3.13: Residues at the site-2 interface potentially affected by pH changes. (a-b) SigmaA-weighted $2F_o - F_C$ electron density map of the eRGMB-NEO1_{FN56} model with the final model shown in stick representation; showing the interactions and coordinated water molecules located at the interfaces between the eRGMB $\beta 5$ - $\beta 6$ loop (green) and the NEO1 FN5 domain (yellow) (b), and the eRGMB $\beta 10$ - $\beta 11$ loop (green) and the NEO1 FN6 domain (yellow) (a). Hydrogen bonds are shown as dotted lines. Charged residues that might be affected by pH changes are labelled. Water molecules are modelled as red crosses.

To test this hypothesis, the SV-AUC experiments with the wild type proteins eRGMB and NEO1_{FN56} were repeated at pH 4.5. Similar to observations at pH 7.0, the resulting distributions for the individual proteins could be fit with single Gaussian curves suggesting that the proteins were monomeric (Fig. 3.14a). However, in contrast to the experiments performed at pH 7.0, the distribution for the eRGMB-NEO1_{FN56} complex at pH 4.5 required only a single Gaussian for fitting (Fig. 3.14b). This peak has an s-value of approximately 3.3, which is higher than that for eRGMB alone. Furthermore, no peaks were observed corresponding to lower molecular weight single components, thereby supporting the assignment of this peak as corresponding to the 1:1 complex. The shoulder on the curve, towards higher s-values, that was indicative of the 2:2 complex in our analysis at pH 7.0 was also not observed at pH 4.5 (Fig. 3.12b). This data appears to support observations from the crystal structures that only a neutral pH favours the formation of the 2:2 complex and suggests that a decrease in pH would likely disrupt this higher order complex.

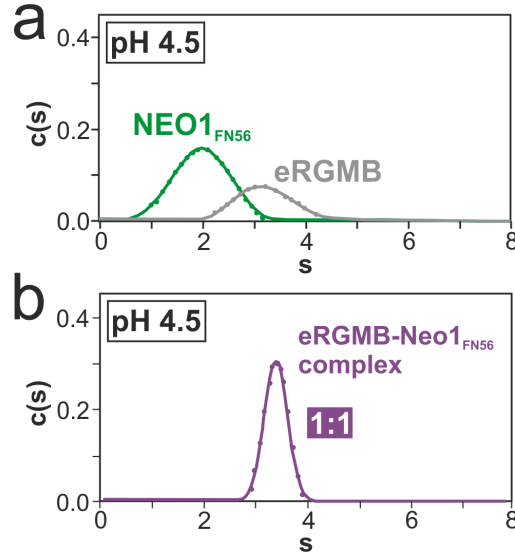


Figure 3.14: The formation of the 2:2 complex is pH dependent. From [240], [Science](#), reprinted with permission from AAAS. (a-b) Analytical ultracentrifugation sedimentation velocity experiments of eRGMB (a, grey) and NEO1_{FN56} (a, green) and the eRGMB-NEO1_{FN56} complex (b, purple) at pH 4.5. Distribution plots obtained from the fitting of SV data using a continuous $c(s)$ distribution model (solid line Gaussian fit to data plotted as points) are shown. Individual Gaussian peaks contributing to the overall distributions are displayed in dotted lines for the eRGMB (a, rmsd = 0.0016), NEO1_{FN56} (a, RMSD: 0.0023) and the eRGMB-NEO1_{FN56} (b, rmsd. = 0.0027) complexes. The individual components run as discrete single, likely monomeric species. The eRGMB-NEO1_{FN56} complex shows a single species most likely corresponding to the 1:1 complex.

3.3.7 The 2:2 eRGMB-NEO1_{FN56} complex is important for signalling

Given that the 2:2 eRGMB-NEO1_{FN56} complex could be observed in solution, it remained to investigate its physiological relevance. In order to test this, wild-type eRGMB, the site-1 mutant eRGMB^{A186R}, which abolishes the eRGM-NEO1_{FN56} 1:1 interaction[240]; and the site-2 mutant eRGMB^{P206N}, which disrupts the formation of the 2:2 complex were expressed and purified. It is well established that binding of RGMs to NEO1 receptors in cerebellar granule neurons (CGNs) *in vitro* triggers growth cone collapse and inhibits neurite outgrowth[85]. In collaboration with the Pasterkamp laboratory the eRGMB mutant proteins were

used in a neurite outgrowth inhibition assay to dissect the roles of the 1:1 and the 2:2 complexes in RGM signalling through NEO1.

Towards this, postnatal mouse CGN explants were cultured on substrates of control (n=23), eRGMB^{A186R} (n=26), eRGMB^{P206N} (n=24) and eRGMB^{WT} (n=27) proteins. As expected eRGMB^{WT} significantly inhibited neurite outgrowth compared to the control substrate (Fig. 3.15, $P<0.01$ and $P<0.05$). This was quantified by measuring the length of the sprouting neurites and comparing the different frequencies of long, medium and short axons (Fig. 3.15b) and the average length of the neurites (Fig. 3.15d). This inhibitory effect was dramatically reduced ($P<0.01$ and $P<0.05$) for eRGMB^{A186R}, where longer neurons were observed with a significantly higher frequency. In addition, eRGMB^{P206N} also exhibited significantly reduced inhibitory activity ($P<0.01$ and $P<0.05$) compared to wild-type eRGMB. Substantial neurite outgrowth was still observed, similar to the axon growth on the control substrate (Fig. 3.15a,c). These results demonstrate that the binding of RGMB to NEO1 through the site-1 interface is required to trigger growth cone collapse. Interestingly, the site-2 interaction also appears to be necessary for full activation of RGM signalling through NEO1, as disruption of this interface inhibited growth cone collapse; suggesting that the 2:2 complex is important in a cellular context.

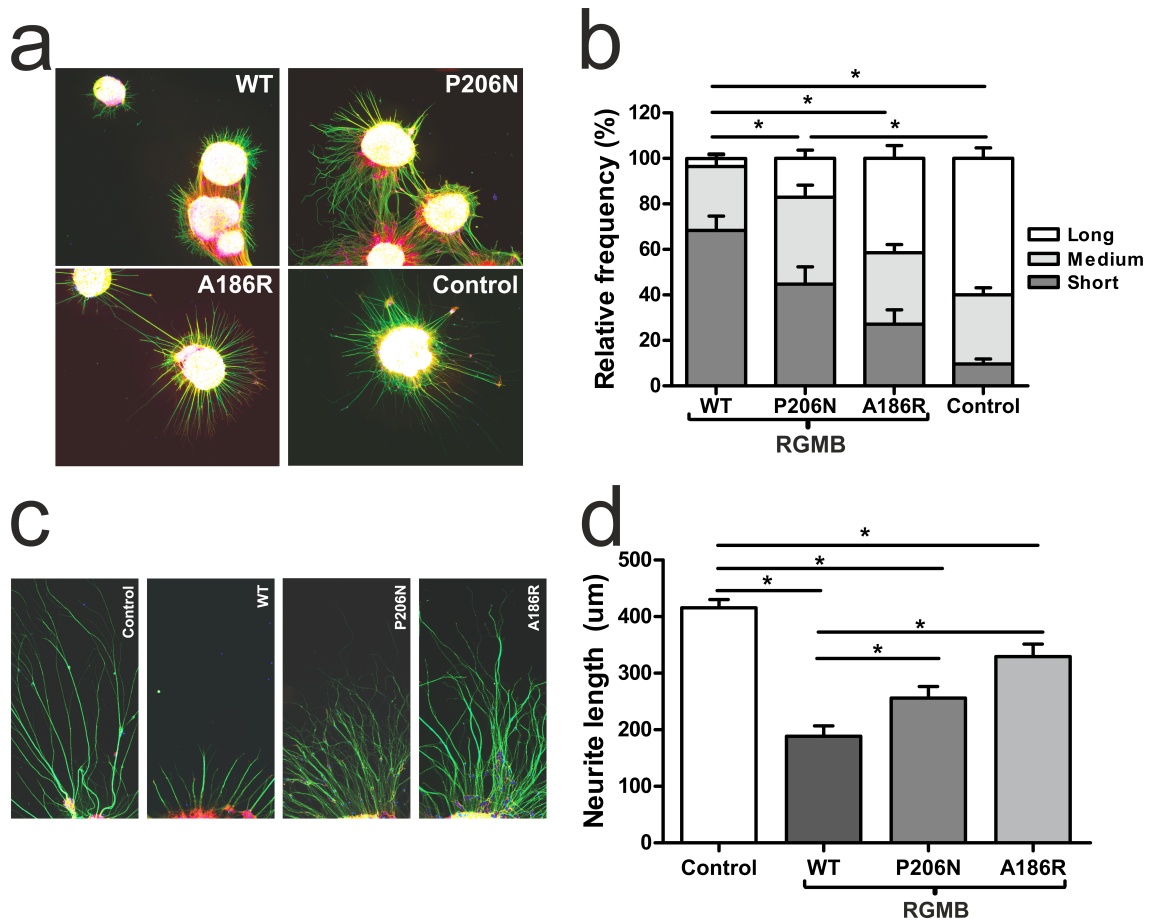


Figure 3.15: Functional analysis of site 1 and 2 mutations on RGM neurite growth inhibitory effects. From [240], [Science](#), reprinted with permission from AAAS. **(a)** Representative P9 mouse CGN explants on coverslips coated with eRGM^{WT} (top left), eRGM^{P206N} (top right), eRGM^{A186R} (bottom left), or control protein (bottom right). Green: β III-tubulin, red: F-actin, blue: nuclei. **(b)** Quantification of CGN neurite outgrowth. Distribution of neurite length (short, medium and long) relative to the control is displayed (number of explants: WT $n=27$, P206N $n=24$, A186R $n=26$, control $n=23$), error bars are s.e.m., * $P<0.01$. **(c)** Higher magnification of representative neuronal explants used in the analysis for (d). The length of individual, non-fasciculated neurites was determined by measuring the distance between the edge of the explant and the growth cone of >20 neurites per explant. **(d)** Average neurite length per explant was calculated for each condition (three independent experiments, number of explants per condition; WT $n=27$, P206N $n=24$, A186R $n=26$, control $n=23$), error bars are s.e.m., * $P<0.05$. (Experiments and analysis performed by Susan van Erp in the laboratory of R. Jeroen Pasterkamp). (Figure adapted from [240]).

3.3.8 NEO1 pre-clusters in the plasma membrane

To investigate the interactions of RGMb and NEO1 in a cellular context using the full-length proteins, co-immunoprecipitation experiments were performed. HEK293T cells were co-transfected with expression plasmids for differentially tagged full-length NEO1 constructs, NEO1-His₆ and NEO1-1D4, and either control plasmid, or plasmid for expression of His₆-tagged, full-length GPI-anchored RGMb constructs RGMb^{WT}, RGMb^{A186R} (site-1 mutant) or RGMb^{P206N} (site-2 mutant). After 40 hours, cell pellets were collected and solubilised in 1.5 % DDM detergent. The concentration of DDM was subsequently lowered to 0.03 % and protein complexes were isolated using 1D4 beads and separated by SDS-PAGE followed by Western blotting with an anti-His₆ antibody. Full-length NEO1 was able to co-immunoprecipitate both wild-type RGMb and RGMb^{P206N} but not RGMb^{A186R} (Fig. 3.16a). This is in agreement with the data showing that RGMb and NEO1 bind directly through the site-1 interface, an interaction that is disrupted by the A186R^{RGMb} mutation. In the presence of RGMs NEO1-1D4 was able to co-immunoprecipitate NEO1-His₆, and although this might have been expected to be abolished for RGMb^{P206N}, as the site-2 interface stabilises the NEO1 dimer interaction in the crystal structure, we did not observe marked differences (Fig. 3.16a). Furthermore, even in the absence of RGMb, NEO1-1D4 was able to co-immunoprecipitate NEO1-His₆ (Fig. 3.16b). This indicates that the NEO1 receptors formed high affinity oligomers in the cellular lysate and suggests that NEO1 may assemble in a pre-clustered form, independently of RGMs, at the cell membrane.

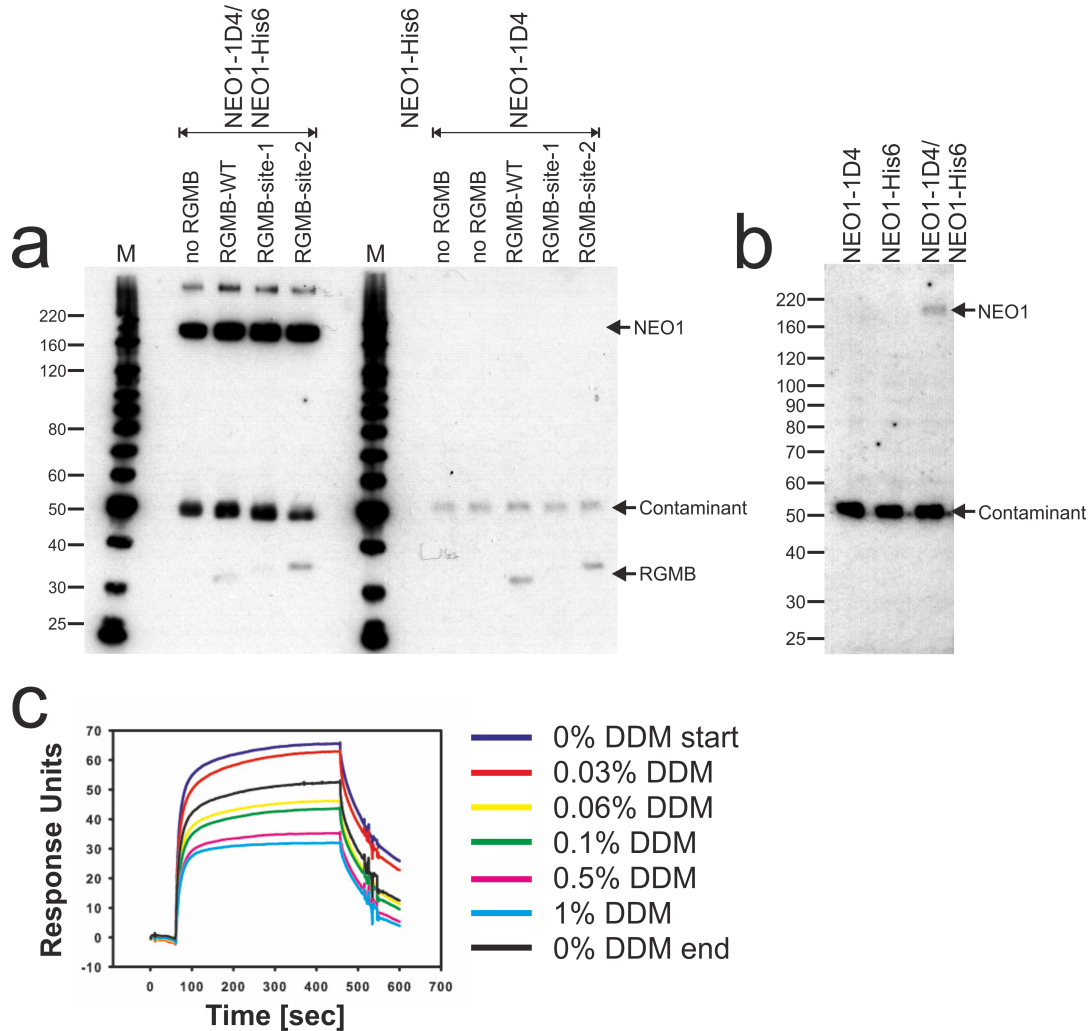


Figure 3.16: Pre-formed NEO1 dimers can be isolated from the cell membrane. From [240], [Science](#), reprinted with permission from AAAS. (a-b) Western blotting analysis of protein complexes eluted from 1D4-antibody-conjugated sepharose beads after pull-down from solubilised cells that were co-transfected with the indicated constructs. All eRGMb constructs were His₆-tagged. The eluted protein complexes were separated by SDS-PAGE (4-12 % Bis Tris gradient gel), and analysed by Western blotting with a mouse anti-His₆ antibody probe. (c) SPR experiments investigating the binding of eRGMb (immobilised on chip) to NEO1_{FN56} (analyte) in different indicated concentrations of Dodecyl maltoside detergent. Binding sensorgrams are shown. Chip concentration was 120 response units. (Figure adapted from [240]).

The amount of RGMB pulled down in the experiment, in comparison to the levels of NEO1 observed on the blot, was lower than expected considering the tight nanomolar interaction between RGMB and NEO1. However, this could have been due to the concentration of DDM that was required throughout the experiment for solubilisation of the full-length proteins. To test this, an SPR experiment was performed to analyse the binding of eRGMB to NEO1_{FN56} across a range of DDM concentrations (Fig. 3.16c). The binding level was somewhat reduced in the presence of 0.03 % DDM (red curve) compared to the binding level in the absence of DDM (blue curve), and a dramatic reduction in binding was observed after treatment with 1 % DDM (cyan curve). Crucially, the level of binding did not recover to that measured prior to addition of the detergent, even after DDM was removed (black curve), suggesting that the high level of DDM we used in the cell lysis may have irreversibly disrupted the RGMB-NEO1 interaction, thereby accounting for the low levels of RGMB in the pull-down.

3.3.9 RGM binding rigidifies the structure of the NEO1 receptor

Having shown that NEO1 receptors can cluster in the membrane independently of RGMs the precise function of RGM binding had still not been ascertained. Towards this, a superposition of all of the available NEO1_{FN56} structures was assembled with respect to the FN6 domain (Fig. 3.17). This included three NEO1_{FN56} apo structures (7 molecules, green shades, cyan and yellow)[240, 148], a NEO1_{FN56}-sucrose octasulphate (SOS) complex (2 molecules, red shades)[240], and the three eRGMB-NEO1_{FN56} complex structures (comprising a total of 5 independent eRGMB-NEO1_{FN56} site-1 interfaces, blue shades and violet)[240]. The eRGMB-NEO1_{FN56} complexes all yielded a very good overlay of the NEO1, with the orientation between the FN5 and FN6 domains showing little varia-

tion and the L3 loop occupying a fixed position (Fig. 3.17). In contrast, the NEO1_{FN56} apo structures show a great deal of flexibility, with up to a 17° angle in the relative orientation of domains FN5 and FN6 and the L3 loop exhibiting varying levels of disorder (Fig. 3.17). This suggests that the binding of RGMB adds rigidity to the conformation of the NEO1 FN5-FN6 interface perhaps by fixing the orientation of the L3 loop.

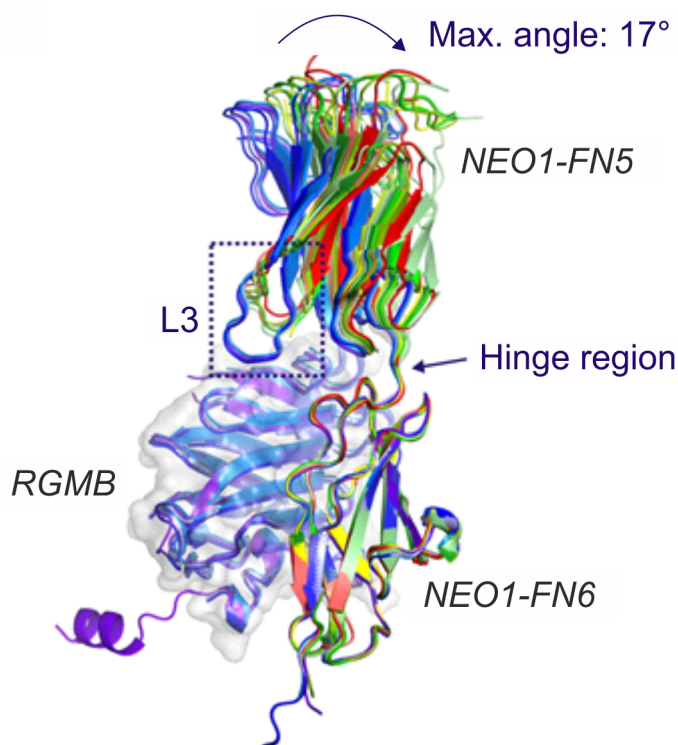


Figure 3.17: NEO1 structure comparison of apo and complex-bound forms. From [240], [Science](#), reprinted with permission from AAAS. Superpositions calculated based on the NEO1-FN6 domain. NEO1_{FN56} apo structures including 4 molecules coloured in different green shades, 2 molecules coloured in different cyan shades (PDB IDs: 4BQ9, 4BQB)[240], and 1 molecule coloured in yellow (PDB ID: 3P4L)[148]; NEO1_{FN56}-sucrose octasulphate (SOS) including 2 molecules, coloured in different red shades (PDB ID: 4BQC)[240]; and three eRGM-NEO1_{FN56} complexes including 2 molecules coloured in dark blue, 2 molecules coloured in light blue and 1 molecule coloured in light violet (PDB IDs: 4BQ6, 4BQ7, 4BQ8)[240]. (Figure adapted from[240]).

3.4 Discussion

3.4.1 A potential model for RGM signalling through NEO1

Taken together these data allow a potential model for RGM signalling through the receptor NEO1 to be proposed (Fig. 3.18). NEO1 is a target for multiple axon guidance cues including the attractive cue Netrin-1, which is known to functionally compete with RGM, suppressing growth cone collapse in dorsal root ganglion axons[153, 141]. The binding site for Netrin-1 on NEO1 has been localised to the FN4-6 domains[146, 147]. The binding of Netrin-1 can be incorporated into the model for RGM signalling and thereby suggest a potential mechanism by which it may inhibit RGM-induced growth cone collapse. Co-immunoprecipitation experiments showed that NEO1 receptors pre-cluster in the membrane, most likely in a loosely defined orientation, with a large degree of flexibility in the juxtamembrane region, as can be inferred from the analysis of NEO1 apo and complex crystal structures (Fig. 3.17). Two RGM molecules act as molecular staples binding to the FN5 and FN6 domains of two NEO1 receptors, fixing the orientation of the juxtamembrane regions of the receptors and stabilising the 2:2 complex, which was shown to be required for fully-active NEO1 signalling (Fig. 3.15). The binding of Netrin-1 to the FN4-6 domains of NEO1 may disrupt the formation of the 2:2 RGM-NEO1 complex, potentially without disrupting the 1:1 binding between RGM and NEO1. RGM can exist in both a membrane tethered and a soluble form after cleavage by proteases[66]. Potentially both forms would be capable of binding NEO1 and activating downstream signalling according to our model. However, it is more likely that the soluble form of RGM plays the larger role as this would allow for long-range actions and for concentration gradients to be established, both important for effective axon guidance.

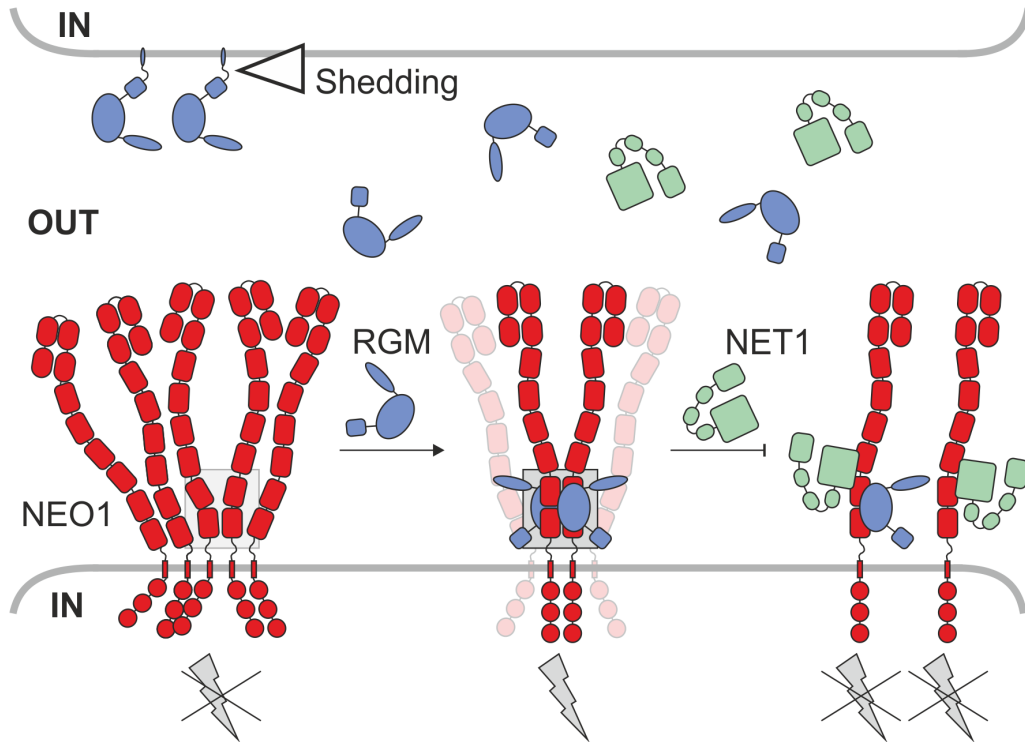


Figure 3.18: Model of trans RGM-NEO1 signalling. From [240], [Science](#), reprinted with permission from AAAS. RGM ectodomains can be shed by proteolytic or phospholipase activity (open triangle). RGM-binding to pre-clustered NEO1 results in formation of NEO1 dimers with a defined, signalling-compatible orientation, that may be part of a supra-molecular clustered state. This arrangement leads to activation of downstream signalling via RhoA (grey lightning bolt). NET1 can inhibit RGM signalling by either simultaneous NEO1-binding or competing with the RGM-NEO1 interaction. The grey box highlights the RGM-NEO1 signalling hub observed in the crystal structure. (Figure taken from [240]).

3.4.2 A pH-dependent mechanism of signalling regulation?

In addition to demonstrating the existence of the 2:2 eRGMB-NEO1_{FN56} complex in solution and validating the interface observed in the crystal structure, AUC experiments also showed that the formation of the 2:2 complex is pH-sensitive and only forms at neutral pH (Fig. 3.14); which may have implications for RGM-NEO1 signalling. The pH at the cell surface is approximately pH 7.5, which would support the formation of NEO1 dimers stabilised by RGMs at the cell-surface. If, however, the complex were to be internalised into endosomes then the pH within these compartments can be as low as pH 5.0, which according to the AUC analysis, is likely to cause dissociation of the 2:2 complex. Targeting the RGM-NEO1 complex for internalisation could function as an additional level of signalling control for efficient termination of the signal. Such a mechanism might ensure that the signal remains short-term and create a system that is highly responsive to changes in RGM levels. There is some evidence of NEO1 internalisation for degradation in the lysosome, which could be linked to this process[284]. It would be interesting to track the subcellular localisation of NEO1 receptors immediately after activation of signalling by RGMs in order to test this hypothesis.

3.4.3 The role of the site-2 interface

The neurite outgrowth inhibition assays showed that the RGMB-NEO1 2:2 complex is important for fully activated signalling (Fig. 3.15). However, subsequent co-immunoprecipitation experiments showed that NEO1 pre-clusters in the membrane (Fig. 3.16), an interaction that is likely due to either the transmembrane or cytoplasmic regions of the receptor, as in the AUC analysis NEO1_{FN56} dimers

were not detected suggesting that the NEO1 dimer interface observed in the crystal structure is not sufficient for receptor dimerisation in the absence of RGMs. This then raises the question of what precisely is the role of RGM in forming this 2:2 complex and why does it enhance signalling if NEO1 is already clustered? Analysis of the eRGMB-NEO1_{FN56} complex structure appears to show that RGM binding to NEO1 rigidifies the FN domains, fixing their relative orientations (Fig. 3.17). Interestingly, the site-1 binding mode appears to be sufficient for this effect and yet both site-1 and site-2 interactions are required for signalling. The site-2 interface is largely hydrophobic in nature with a relatively modest buried surface area (695 Å²), markedly lower than the site-1 interface area (2350 Å²) and slightly smaller than the NEO1 dimer interface (835 Å²). However, it appears to be a relatively robust interface as the mutagenesis and AUC experiments demonstrated that the 2:2 complex was not disrupted by the V204G/P206G double mutation. Although, this was quite a conservative mutation as the V204 hydrogen bond forms through its backbone carbonyl and as such the glycine substitution would not necessarily disrupt this, and the proline to glycine substitution does not introduce steric hindrance or charge affects. It may be that the site-2 interactions of the RGM molecules with the NEO1 receptors staple the relative orientations of each NEO1 monomer in a signalling-competent arrangement. It would be interesting to test the V204G/P206G mutant in the neurite growth inhibition assay as, even though it does not inhibit 2:2 complex formation, the mutation may not facilitate the formation of the rigid structure of the complex that is necessary for signal transduction.

3.4.4 NEO1 as a central target for RGMs and Netrin-1

A recent crystal structure has shed light on the binding mode between Netrin-1 and NEO1[147]. Netrin-1 comprises an N-terminal laminin domain (LN), fol-

lowed by three cysteine-rich LN-type epidermal growth factor (EGF)-like modules (LE1, LE2, and LE3), and a small positively charged C-terminal domain (LC). The complex crystal structure describes a Netrin-1 dimer in an X-shape linking two NEO1 molecules. Each Netrin-1 LN domain contacts a NEO1 FN4 domain and each Netrin-1 LE3 domain binds to a NEO1 FN5 domain[147]. The Netrin-1 molecules form a dimer interface between their LE2 domains. Interestingly in this complex the C-termini of the two NEO1 receptors are approximately 105 Å apart, in contrast to the eRGMB-NEO1_{FN56} structure in which they are less than 60 Å apart. This potentially suggests that the attractive guidance role attributed to Netrin-1 binding to NEO1 is simply an effect of disrupting NEO1 signalling, which activates downstream pathways triggering growth cone collapse. The disruption of NEO1 dimerisation explains observations that Netrin-1 inhibits RGM-mediated growth cone collapse. The non-overlapping binding sites of RGM and Netrin-1 on NEO1 suggest that they could bind the receptor simultaneously. It would be interesting to investigate whether this triple complex can assemble in solution and then to test the affects of both proteins alone and in combination in neurite outgrowth assays.

3.4.5 Insights into the JHH-causing mechanisms of RGMC mutations

Twelve out of the 14 JHH-linked mutations in RGMC abolish the secretion of the protein (Fig. 3.7). Given that seven of these mutations cluster around the autocatalytic cleavage and three more map onto the β -sandwich it is likely that they affect the stability of the RGMC middle domain thereby impairing protein secretion. RGMC-activated BMP6 signalling is the mechanism by which levels of the peptide hormone hepcidin are regulated, which is essential for iron homeostasis[107, 102]. The absence of RGMC triggers extreme iron overload,

as was exhibited in the livers of RGM^{C-/-} knockout mice[106]. This analysis therefore provides a molecular rationale for JHH-linked mutations in the RGM middle domain: if RGM is not secreted then it cannot enhance BMP signalling, causing downregulation of hepcidin levels and resulting in JHH. Interestingly, two mutations G99R and L101P, which localise to the N-terminal domain of RGM, were secreted. These must therefore disrupt RGM-mediated activation of BMP6 signalling by another mechanism. In fact, work in Chapter 4 will discuss these mutations and a potential disease-causing mechanism.

3.5 Summary

This work has shed light on the molecular mechanism of RGM signalling through NEO1. A 2:2 form of the RGM-NEO1 complex, originally observed in a crystal structure, was demonstrated not only to exist in solution but also to be physiologically important. An analysis of the eRGM-NEO1_{FN56} structures has demonstrated that RGMs fix the relative orientations of the FN5 and FN6 domains of NEO1 within the architecture of the 2:2 complex and that this converts inactive NEO1 clusters into signalling competent arrays. Furthermore, this work has demonstrated a potential mechanism behind JHH-causing mutations in RGM. It is interesting to consider that given the involvement of RGMs in the BMP signalling pathway and the role of NEO1 in Netrin-1 signalling that the RGM-NEO1 complex may act as a signalling hub, serving as a nucleation centre for a multitude of guidance and morphogenic cues.

Chapter 4

Molecular mechanisms of RGM-activated BMP signalling

4.1 Abstract

RGMs are known to be important activators of BMP signalling, a process that is crucial in the regulation of iron metabolism, but has also been linked to many other processes including the expression of IL-6 in the immune response, and kidney and brain development. However, although RGMs are known to bind directly to the BMP ligand, the molecular mechanism of RGM-activated BMP signalling has yet to be determined. In this chapter, I present crystal structures of all RGM family members in complex with the BMP2 ligand, which demonstrate that the RGM N-terminal domain is the site of BMP2 interaction and provide a molecular rationale for JHH-causing mutations in RGMC. I describe experiments demonstrating that RGM and BMPR1A compete for binding to the BMP2 ligand and that the two proteins have different effects on BMP signalling in luciferase reporter assays. Furthermore, I present MALS data showing that unlike the BMPR1A-BMP2 complex, the RGM-BMP2 complex is pH-sensitive. Taken together, these data allow us to propose a model for RGM-activated BMP signalling, which is linked to endocytosis and related to the subcellular localisation of the BMP receptors.

4.2 Introduction

RGMs have been linked to a diverse catalogue of biological functions and disease states, not all of which can be mediated by signalling through their canonical cell surface receptor NEO1. All RGMs have also been identified as important co-receptors for BMP signalling and for this mode of signalling there is currently no structural information describing RGM interactions with any components of the BMP pathway. The architecture of the activating signalling complex has not been identified and as such, the precise mechanism of RGM-mediated activation

of BMP signalling is not clear.

4.2.1 Structural studies of BMP signalling

Aside from the lack of insight into the involvement of RGMs, there is a wealth of structural data describing the BMP family, including high-resolution structures for many of the BMP ligands and the ectodomains of the type I and type II receptors. BMP ligands are expressed with a large pro-domain that is required for protein folding and is then cleaved to yield the mature protein. Each BMP monomer consists of a well-characterised cysteine knot domain stabilised by three disulphide bonds. The active form of the BMP ligand is a constitutive dimer, with the monomers linked by an intermolecular disulphide bond (Fig. 4.1a). The overall topology of the fold of the BMP ligand, shown here for BMP2[168], has been described as a hand[285], with the helix $\alpha 3$ as the wrist, the cysteine knot as the palm and the extended β -sheets as the fingers (Figs 4.1a,b). This fold is highly conserved across BMPs and TGF β s, for example the only structural element unique to BMP2 is the small β -sheet $\beta 5a$; in addition, the α -helices $\alpha 1$ and $\alpha 2$ are unique to TGF β s and $\alpha 4$ is only found in BMP7[286] (Fig. 4.1b). The interactions between the $\alpha 3$ helices and the β -sheets of the opposite monomer create two stable hydrophobic cores in the protein[168]. This, combined with the intermolecular disulphide, make the BMP dimer unusually stable and explains the use of strikingly harsh buffers in the dimer purification protocols[287, 288] (see Materials and Methods).

Examination of the arrangement of the BMP molecules in the dimer demonstrates that the monomers assemble at approximately a 40° angle[168]. This results in numerous cavities, which were immediately realised to have potential as interaction sites (Fig. 4.1c). The so-called “cavity I” comprises many positively charged residues[168] and has been identified as a heparin binding

site[289]. The basic residues from the N-terminal region of the protein are presumed to contribute to this binding site, however this region is disordered in all BMP ligand structures, although in structures of TGF β it forms a helix[285]. The function of “cavity II” has not been identified. The remaining “finger-helix” symmetry-related cavity was later identified as the type I receptor binding site.

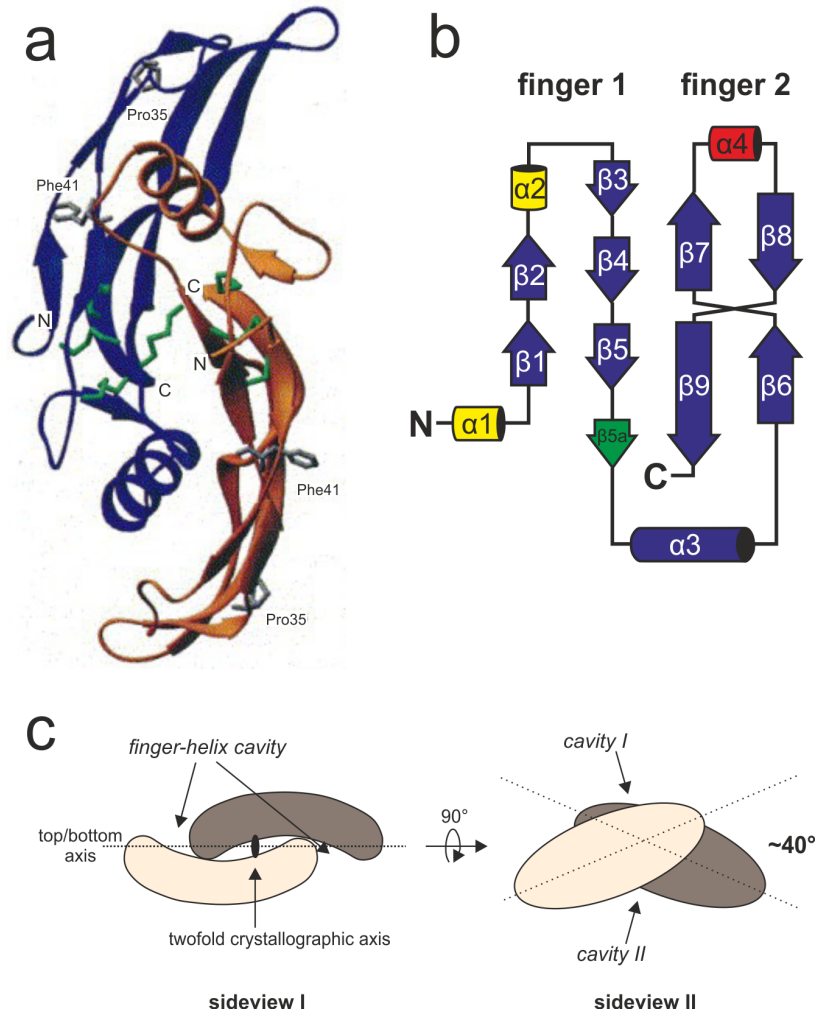


Figure 4.1: Structure of the BMP2 dimer. (a) The BMP2 dimer structure shown in cartoon form. BMP2 monomers shown in blue/orange and disulphide bonds shown as green sticks. Residues with irregular conformations are labelled. The N- and C-termini are indicated. Image reprinted from [Journal of Molecular Biology](#) [168], Copyright 1999, with permission from Elsevier. (b) Topology diagram of BMP2, helices represented as cylinders and β -strands as arrows. Secondary structural elements conserved across the TGF β /BMP family are shown in blue, those unique to TGF β in yellow, to BMP7 in red and to BMP2 in green. (c) Schematic showing the relative orientation of the BMP2 monomers within the dimer. The angle between them $\sim 40^\circ$; cavities, identified as potential protein binding sites, are labelled. Figures from[168].

BMP dimers signal through binding type I and type II receptors. The type I receptor is the high affinity receptor for the BMP family, excluding BMP7 which retains the TGF β -associated specificity for high affinity type II receptor binding[290, 291, 292, 293]. The crystal structure of the BMP2-eBMPR1A receptor complex revealed the structure of the ectodomain of the receptor and its mode of interaction with BMP2[241] (Fig. 4.2a). The BMPR1A ectodomain forms a compact single-domain structure stabilised by disulphide bonds comprising two β -sheets (made up of strands β 1-2 and β 3-5). Two eBMPR1A molecules bind at the finger-helix cavities of BMP2 and make contact with both BMP molecules in the dimer. Interestingly the surface complementarity between eBMPR1A and BMP2 in the complex is not perfect and in many cases there is space for larger side-chains. This illustrates the potential for promiscuity between receptors and ligands in the BMP signalling system and explains why different ligand-receptor combinations have different affinities[294] (for review[295]). The interface between eBMPR1A and BMP2 is comprised largely of hydrophobic contacts. An interaction identified to be of particular importance is that of F108^{BMPR1A}, which extends from helix α 1 into a pocket on BMP2 where it is surrounded by numerous hydrophobic residues including W310^{BMP2} and W313^{BMP2} (Fig. 4.2b). Not only are these pocket residues highly conserved within the TGF β superfamily[168] but all type I receptors (except ALK-1) feature a large hydrophobic residue at the site corresponding to F108^{BMPR1A} ([241]). This conserved helix α 1-hydrophobic pocket motif is important for specifying type I receptor binding at the finger-helix cavity. Following the structure solution of the BMPR1A-BMP interaction, the structure of a BMP ligand in complex with a type II receptor was also solved[241]. The structure of the BMP7-eACVR2 complex showed that the concave surface of the type II receptor binds at the convex face of the BMP finger region[296] (Fig. 4.2c), a mode of interaction that had been predicted from prior structural information[297, 241]. The ectodomain of ACVR2A is made up largely of β -sheets and is again stabilised by several disulphide bonds.

Subsequently, the structure of the BMP signalling complex was solved. It featured one BMP2 dimer bound to four receptor ectodomain molecules including two copies of ACVR2A and two copies of BMPR1A[298] (Fig. 4.2d). In the complex the receptor ectodomains do not contact each other meaning that any allosteric interactions must be propagated through the ligand; in fact BMP receptor binding has been shown to be uncooperative[291, 292]. This is in contrast to TGF β receptor binding, which has been shown to be highly cooperative[290, 299, 293] and receptor-receptor contacts can be observed in the crystal structure[300, 301, 293] (Fig. 4.2e). Experiments revealed that the amino acid combinations at the receptor binding sites on the BMP ligand determine receptor affinities, for example a single point mutation on BMP2 can increase its binding affinity for ACVR2A five fold[298]. The hexaheteromeric architecture of this complex is thought to represent the form of the active BMP signalling complex *in vivo*.

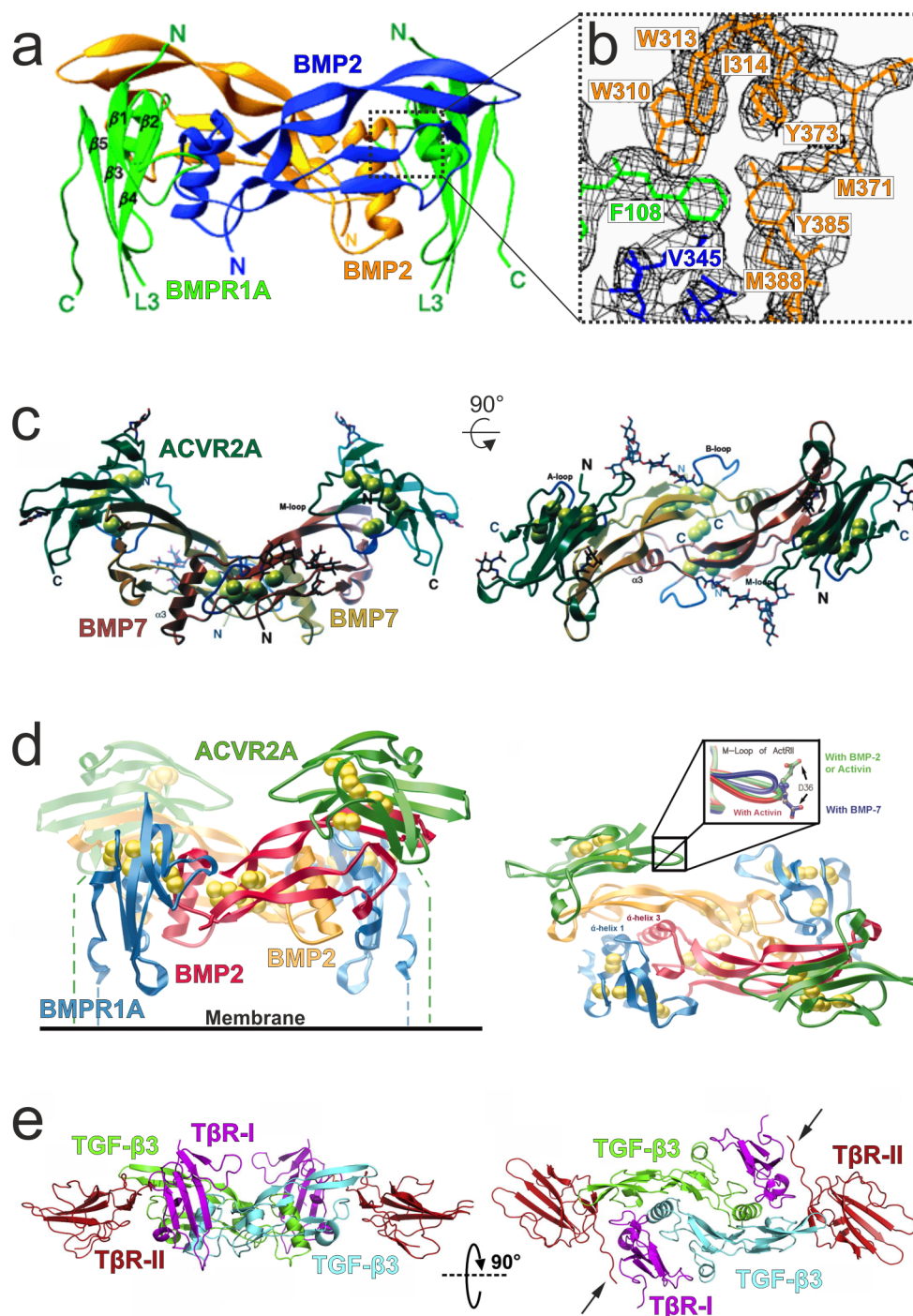


Figure 4.2: Structures of BMP ligand-receptor complexes. Structures shown in cartoon representation. **(a)** The BMP2-eBMPR1A complex structure. BMP2 = blue/orange, eBMPR1A = green. The N- and C- termini, the unresolved L3 loop and the eBMPR1A β -strands are labelled. Image reprinted by permission from Macmillan Publishers Ltd: [Nature Structural and Molecular Biology](#) [241], copyright 2000. **(b)** Close-up view of the BMP2-eBMPR1A interface. F108^{BMPR1A} protrudes into a BMP2 hydrophobic pocket. Image reprinted by permission from Macmillan Publishers Ltd: [Nature Structural and Molecular Biology](#) [241], copyright 2000. **(c)** Structure of the BMP7-eACVR2A complex. BMP7 = red/gold, ACVR2A = dark green, disulphide bonds = gold spheres. Image reprinted from [Molecular Cell](#) [296], Copyright 2003, with permission from Elsevier. **(d)** Structure of the ternary eBMPR1A-BMP2-eACVR2A active signalling complex. eBMPR1A = blue, eACVR2A = green, BMP2 = red/gold, disulphide bonds = yellow spheres. Attachment of the receptors (dotted lines) to the membrane (black line) is shown. Variation in the ACVR2A M-loop, conferring receptor specificity is shown (right panel). Image reprinted from [Proc Natl Acad Sci](#) [298], Copyright 2006. **(e)** The structure of the ternary TGF β active signalling complex. T β R-I = purple, T β R-II = red, TGF- β 3 = cyan/green. Arrows indicate T β R-II-T β R-I interaction. Image reprinted from [Molecular Cell](#) [300], Copyright 2008, with permission from Elsevier.

4.2.2 RGMs as activators of BMP signalling

All mammalian RGMs have been established as important co-receptors for BMP signalling[102, 83, 89]. However, it remains to be determined how RGMs fit into the architecture of the active BMP signalling complex. RGMs were shown to bind directly to the BMP ligand with nanomolar affinities[302]. All RGMs show a preference for binding to BMP4 with the tightest RGM-BMP interaction between RGM-B and BMP4 ($K_d = 2.6$ nM). RGM-C shows a high affinity for BMP6 with a K_d of 8.1 nM, which is expected as this is the endogenous ligand involved in iron regulation[107]. RGM-B shows a particular preference for the BMP2/4 subfamily compared to the BMP5/6/7 subfamily, for example RGM-B demonstrates a weak interaction with BMP7 with a K_d of 166 nM. In addition, none of the RGMs bind to BMP9, which has been shown to use a different co-receptor called Endoglin[303]. Overall there is a range of affinities for the interactions between different RGMs and the different BMP ligands, which may affect the patterns of RGM-mediated BMP signalling in different tissues. Immunoprecipitation data also showed that RGMs bind to all BMP type I receptors and the BMP type II receptors ACVR2A and ACVR2B[83, 89]. There is some evidence suggesting that RGMs enhance the binding of the type I receptor ALK6 to the BMP ligand and form an RGM-BMP-ALK6 complex[102, 83]. Furthermore, it was shown that RGM-A alters the utilisation of BMP type II receptors by BMP2 and BMP4[189]. In mouse arterial smooth muscle cells BMP2/4 signal exclusively through BMPR2 while in BMPR2-knockout cells, these ligands can signal through ACVR2A. A further knockdown of RGM-A using targeted siRNAs abolished the alternative utilisation of ACVR2A by these ligands and inhibited BMP signalling. A model for RGM-A-mediated BMP2/4 signalling was proposed in which a BMPR1A-BMP2/4-RGM-A complex enhances the binding affinity of ACVR2A allowing utilisation of both type II receptors leading to an increased BMP signal[189]. These data suggest that RGM-A is incorporated into the ac-

tive hexaheteromeric BMP signalling complex and would likely make contacts with both receptor ectodomains thereby enhancing their binding affinities for the BMP ligand. The architecture of this super-complex has yet to be determined, and in fact it remains to be seen whether RGMs enhance the binding affinity of BMP ligands for BMP receptors *in vivo* on the cell surface.

4.2.3 JHH-linked mutations may identify the BMP binding site on RGMs

Over fourteen JHH-linked mutations have been described in RGM. Ten of these mutations are located in the middle domain, while twelve abolish secretion of recombinant RGM protein in a mammalian expression system (Fig. 3.7)[240]. This allowed a rationale for the mechanism by which these mutations cause JHH to be proposed: secretion of RGM is inhibited, preventing RGM-mediated BMP6 signalling, which is responsible for up-regulation of hepcidin, thereby resulting in iron accumulation in parenchymal tissues. However, two JHH-linked mutations in RGM are secreted in a mammalian expression system, G99R^{RGM} and L101P^{RGM} (Fig. 3.7). This implies that these mutants must inhibit BMP6 signalling by another mechanism, perhaps by directly affecting the BMP6-RGM interaction. Both of these residues are located in the N-terminal domain of RGM suggesting that this domain may comprise the BMP binding site.

4.2.4 Chapter outline

The aim of this chapter was to test the hypothesis regarding the location of the BMP binding site on RGMs and to solve the crystal structure of the RGM-BMP2 complex. Towards this, expression constructs for the N-terminal domain of all RGM family members were designed and purification strategies for RGM_{ND} and

BMP2 were optimised. The RGM N-terminal domain was found to be sufficient for RGM-binding to BMP2 using SPR and the structures of the N-terminal domains of all RGM family members in complex with BMP2 were solved. These crystal structures guided subsequent mutagenesis and SPR studies to validate the observed binding interface and also to confirm the hypothesis for the disease-causing mechanism of two JHH-linked mutations. In addition, a structure of the BMP2-eBMPR1A complex was solved and revealed previously unseen glycosylation sites and a novel N-terminal BMP2 α -helix. Structural comparisons of the BMP2-eBMPR1A and BMP2-RGM_{ND} complexes, along with competition SPR experiments, revealed that BMPR1A and RGM compete for the same binding site on BMP2. The ectodomains of these two proteins were shown to have different effects on BMP signalling in cellular luciferase reporter assays and the BMP2-RGMB complex was found to be pH-sensitive. These data allowed an endocytosis-linked model for RGM-activated BMP signalling to be proposed. Several fluorescence microscopy experiments were performed with the aim of testing this hypothesis, including tracking the internalisation of fluorescently labelled BMP2 ligand in live cells and FRET-FLIM analysis of BMP receptor interactions. Unfortunately, the results of these challenging experiments were inconclusive and some of the reasons for this are discussed. Finally this chapter concludes with a discussion of the findings in the context of what is already known regarding RGM-activated BMP signalling.

4.3 Results

4.3.1 Expression and purification of RGM_{CND} and BMP2

Expression constructs for the N-terminal domains of the RGMs (RGM_A_{ND}: T45-S139; RGM_B_{ND}: Q53-H136; RGM_C_{ND}: Q36-P147) were designed based on the eRGM_B-NEO1_{FN56} crystal structure and the N-terminal conserved region in sequence alignments of vertebrate RGMs, ensuring an even number of cysteines were included for disulphide bond formation (Fig. 4.3a,b). These constructs were cloned into the pHLsec vector with a C-terminal His₆ tag for expression in mammalian cells as described in Materials and Methods (Section 2.1 and 2.3), shown here for RGM_C_{ND} (Fig. 4.3c). Secreted RGM_C_{ND} was isolated by metal ion affinity and purified by SEC; the protein appeared monodisperse, exhibiting a single peak in the SEC trace and no aggregation was observed (Fig. 4.3c). Following treatment with Endo-F1, to remove flexible N-linked glycans that can hinder crystallisation, RGM_C_{ND} was re-purified by SEC. Analysis of the peak fractions by SDS-PAGE revealed that the protein migrated as several closely spaced bands at around the expected size of 12 kDa, most likely indicating several different glycosylation forms, therefore, only the most homogenous RGM_C_{ND} SEC fractions were pooled for crystallisation (Fig. 4.3c, left panel).

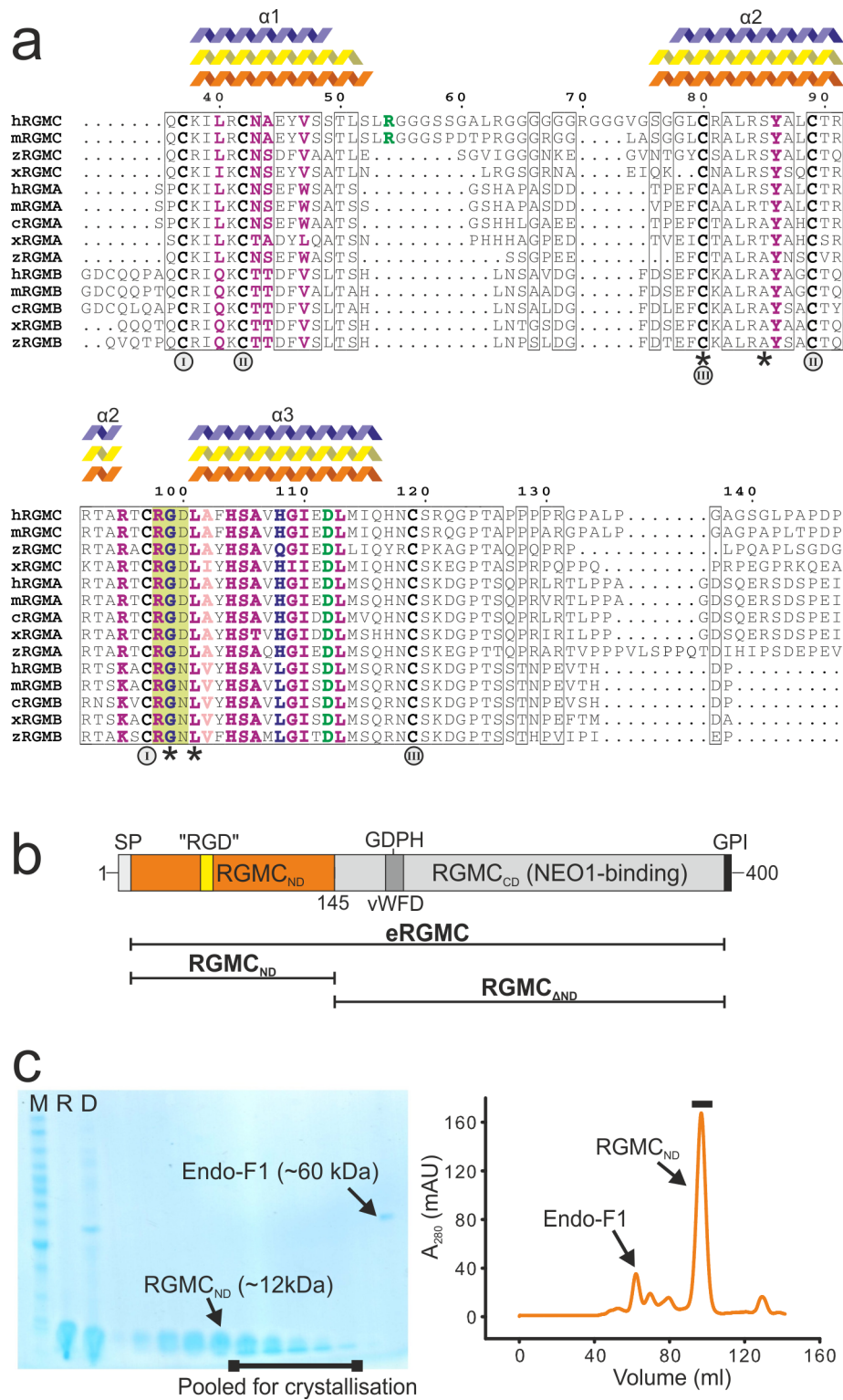


Figure 4.3: Expression and purification of RGM_{CND}. (a) Sequence alignment of the N-terminal domain of RGM family members. h = human, m = mouse, z = zebrafish, x = *Xenopus*. Numbering is full length hRGM. Secondary structure displayed: RGMA_{ND} (blue), RGMB_{ND} (yellow), RGM_{CND} (orange). RGM residues interacting with BMP highlighted: hydrophilic interactions = dark blue, RGMA only = salmon, RGM only = green; non-bonded contacts = magenta. Disulphide bridges labelled with Roman numerals; asterisks (*) mark hRGM disease-linked mutations. (b) Schematic of hRGM, N-terminal domain (RGM_{CND}) = orange; C-terminal region (RGM_{CD}) = grey, SP = signal peptide, GPI = glycosylphosphatidylinositol anchor, vWfD = von Willebrand Factor D-like domain. The “RGD” motif and the autocatalytic cleavage site (“GDPH”) are shown. The construct boundaries for eRGM, RGM_{CND} and RGM_{ΔND} are shown. (c) SDS-PAGE analysis (left panel) of deglycosylated (Endo-F1) RGM_{CND} SEC fractions (black bar, right panel). M = marker, R = RGM_{CND}, D = deglycosylated RGM_{CND}. Calculated mass shown. Pooled fractions labelled.

The mature human BMP2 peptide (Fig. 4.4a,b) was expressed in bacterial inclusion bodies and the protein was refolded according to a protocol outlined in Materials and Methods (Section 2.3.1, adapted from[206]) that biases the formation of the folded disulphide-linked dimeric species (Fig. 4.4c). The BMP2 monomer was first refolded in the presence of a 1:1 GSSG:GSH redox system, then one half of the mixture was incubated with an excess of GSSG before the two reaction halves were recombined resulting in a mixture of BMP2 species, with a majority in dimeric form. Initially BMP2 monomers were removed by heparin affinity chromatography (Fig. 4.4d), following this SEC was performed to separate higher-order oligomers from BMP2 dimers (Fig. 4.4e). The BMP2 dimer migrated at the expected size of approximately 26 kDa and fractions with a purity of greater than 95 % were pooled for activity tests (Fig. 4.5).

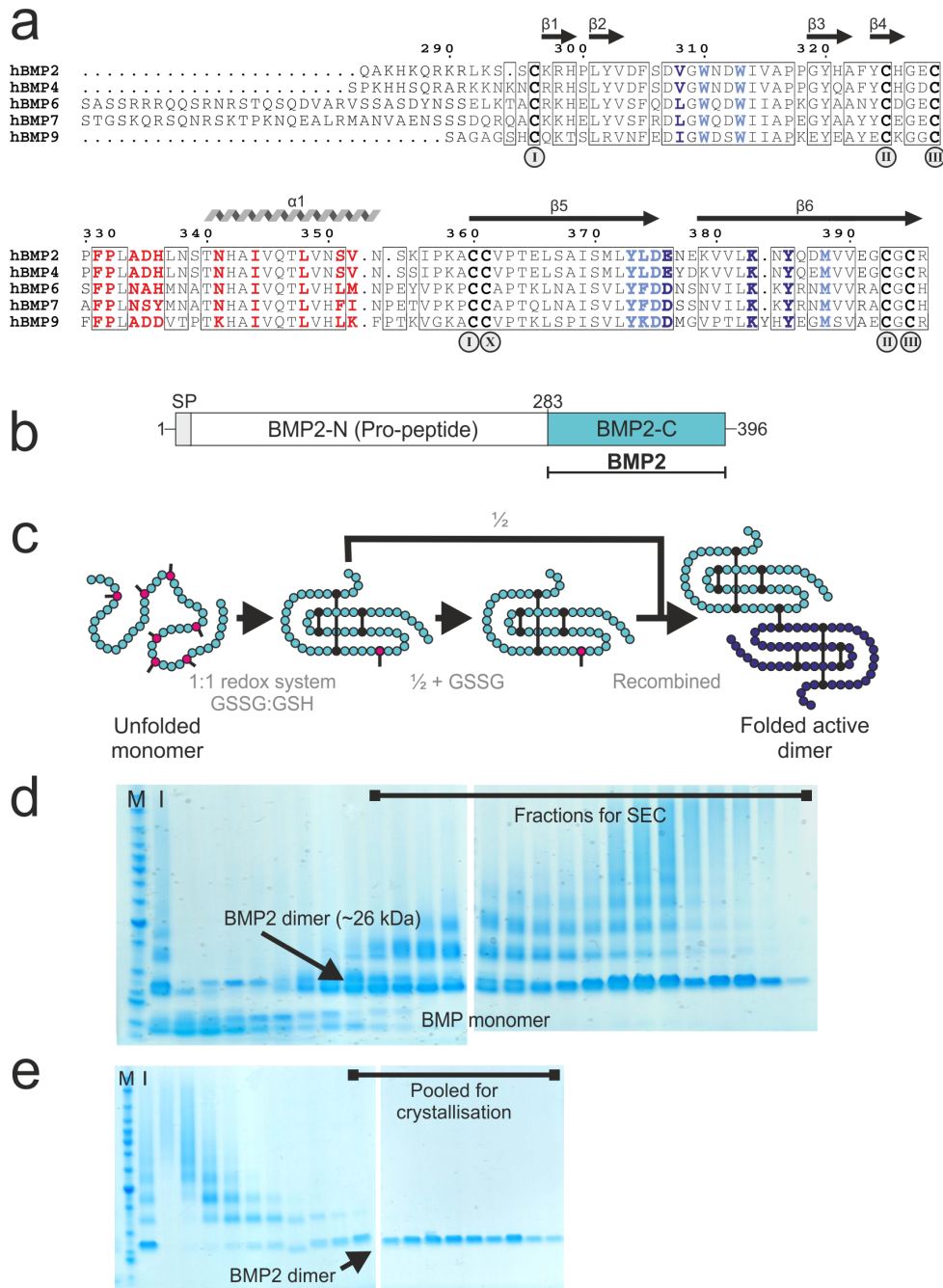


Figure 4.4: BMP2 expression and purification. (a) Sequence alignment of human BMPs. Numbering is full length hBMP2; BMP2 residues interacting with RGM highlighted: molecule 1 hydrophilic interactions = dark blue, non-bonded contacts = light blue; molecule 2 non-bonded contacts = red. Disulphide bridges indicated by Roman numerals; X = intermolecular disulphide. Secondary structure shown above the alignment. (b) Schematic of hBMP2, expression construct labelled. SP = signal peptide. (c) Schematic of the BMP2 dimer refolding scheme (adapted from [206]). GSSG = oxidised L-glutathione, GSH = reduced L-glutathione. (d-e) SDS-PAGE analysis of heparin affinity chromatography (d) and SEC (e) fractions. Pooled fractions are indicated (black bars).

To confirm the activity of the purified, refolded BMP2 dimer, two optimised activity assays were performed (Materials and Methods, section 2.2 and 2.10, for detailed protocols). An anti phospho-SMAD1/5 antibody was used to assay for phosphorylation of the downstream SMAD proteins before and after addition of BMP2 (Fig. 4.5a). This assay demonstrated that the refolded BMP2 dimer was able to trigger activation of the SMAD1/5/8 pathway in three different cell-lines (COS7, C2C12 and HeLa cells). In addition, a stably transfected cellular luciferase reporter assay[304] was used to show that refolded BMP2 could activate expression of the luciferase gene under the control of the Id1 BMP target promoter (Fig. 4.5b). These assays confirmed that the refolded BMP2 dimer was active and therefore correctly folded.

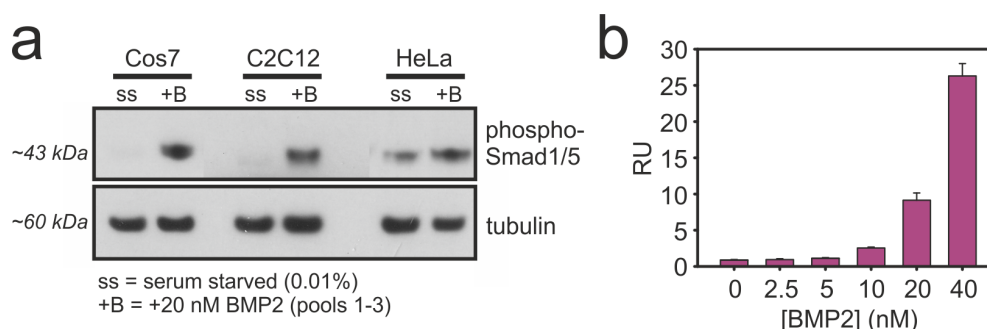


Figure 4.5: BMP2 activity assays. (a) SMAD phosphorylation assay. Western blot using an anti-phospho-SMAD1/5 antibody. An anti-tubulin antibody was used as a loading control. Samples from COS7, C2C12 and HeLa cells are indicated; ss = serum-starved, +B = 20 nM BMP2 added. (b) BRE-luciferase assay in C2C12 cells stably transfected with the BRE-Luc plasmid[304]. RU = relative luciferase response units. Error bars are s.e.m., from eight cell cultures and two independent experiments.

4.3.2 The structure of the RGM_{CND}-BMP2 complex

Towards determining the structure of the RGM_{CND}-BMP2 complex, purified RGM_{CND} and BMP2 were mixed in a 1:1 molar ratio and the complex was concentrated to a final concentration of 5 mg/ml for crystallisation trials. Crystalli-

sation trials were also conducted for the RGMA_{ND} -BMP2 and RGMB_{ND} -BMP2 complexes and in all three cases crystals were obtained (Fig. 4.6). The crystals were treated with cryoprotectant and cryo-cooled in liquid nitrogen before being analysed at Diamond Light Source UK (see Materials and Methods). Crystals of all three RGM_{ND} -BMP2 complexes diffracted to high resolution (Table 1).

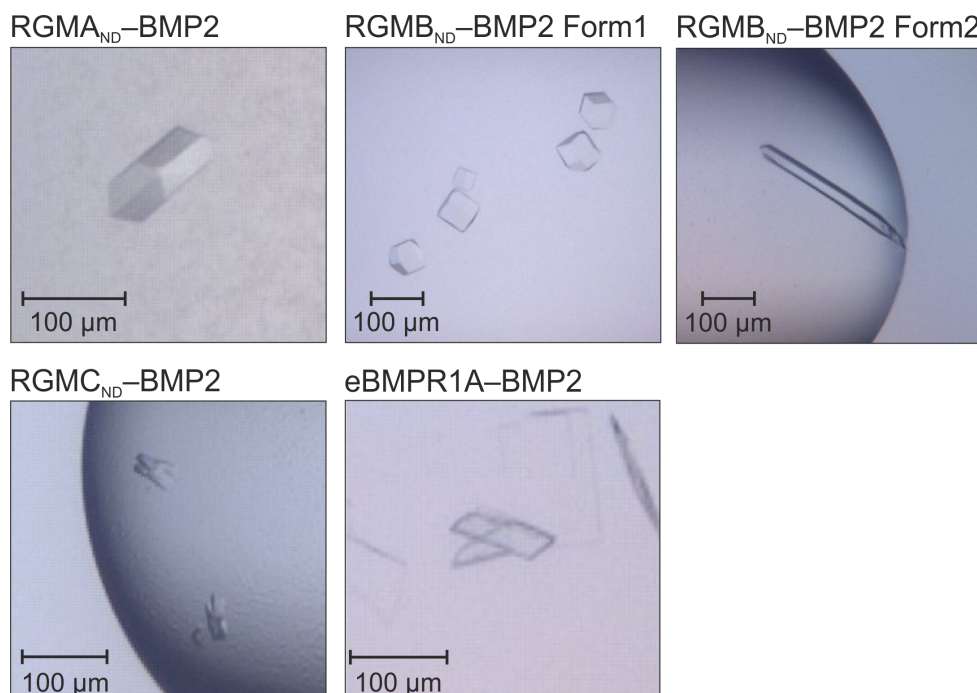


Figure 4.6: Crystallisation of RGM_{ND} -BMP2 and eBMPR1A-BMP2 complexes. Crystallisation images of each protein complex described in this chapter. Images recorded from crystallisation plates maintained in a TAP Home-base storage vault, imaged via a Veeco visualisation system[211].

The crystal structure of BMP2 in complex with RGMC_{ND} was solved to 2.35 Å resolution (Fig. 4.7 and Table 1) and after phasing by molecular replacement using the published BMP2 dimer structure (PDB ID: 3BMP[168]), electron density for RGMC_{ND} was immediately discernible (Fig. 4.7). In the complex, two molecules of RGMC_{ND} bind to one disulphide-linked BMP2 dimer and the two RGMC molecules in the complex are related by a non-crystallographic pseudo-twofold axis and have an r.m.s.d. of 0.73 Å for 66 equivalent $\text{C}\alpha$ positions (Fig. 4.7a and Fig. 4.9a).

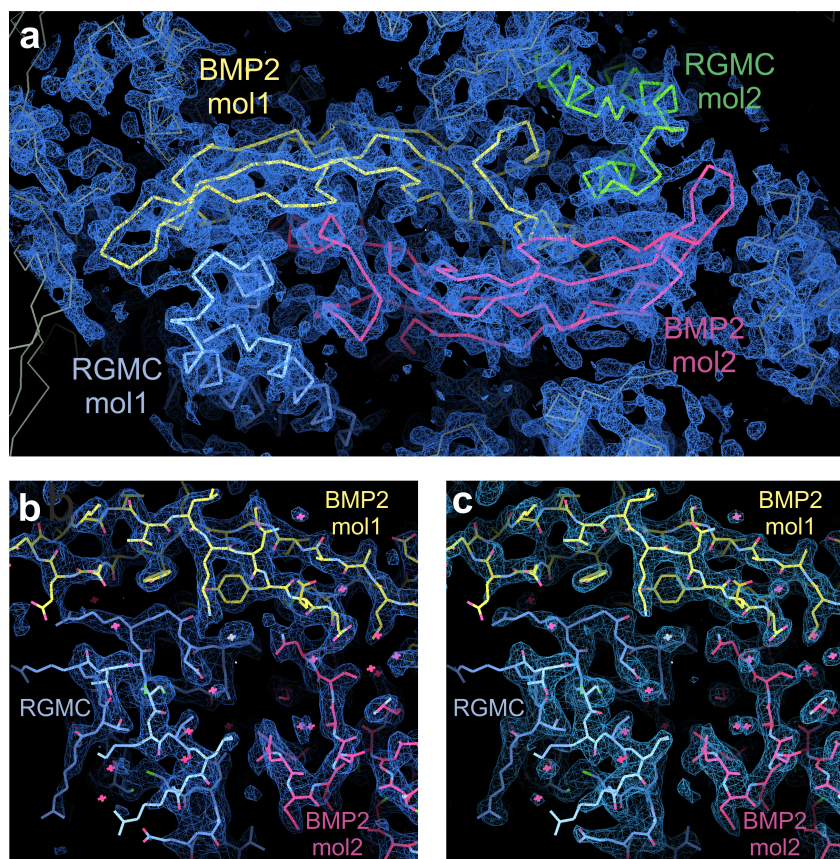


Figure 4.7: Electron density maps of the RGM-C-BMP2 complex. Reprinted by permission from Macmillan Publishers Ltd: [Nature Structural and Molecular Biology](#), [320] copyright (2015) (a) Initial electron density map, at 2.35 Å resolution, of the RGM_{ND}-BMP2 complex after molecular replacement (using only the BMP2 dimer as search model) in PHASER[238] and density modification in PARROT[239]. Contour level is 1.0 σ . The final refined RGM_{ND}-BMP2 model is represented as ribbon (RGM-C: blue and green; BMP2: yellow and magenta). (b-c) Close-up view onto the RGM-C-BMP2 interface. Protein chains are depicted in stick representation. Colour coding is as in (a). (b) Initial electron density map as described in (a). (c) SigmaA-weighted $2F_o - F_c$ map after the final round of refinement from autoBUSTER[247] contoured at 1.0 σ .

RGM_{ND} (with amino acid residues S35-R54 and G76-P129 resolved in the structure) adopts a novel fold composed of a triple helix bundle stabilised by three disulphide bonds (Fig. 4.8a,b and 4.9). The helices pack together forming a hydrophobic core at their centre (Fig. 4.8b). The long flexible loop between helices α_1 and α_2 was not ordered in the crystal structure, whereas the loop between helices α_2 and α_3 was ordered, presumably because its position is fixed by a disulphide bond (Fig. 4.8b,c). The C-terminal tail of RGM_{ND} in the structure

is orientated away from the triple helix bundle by a disulphide bond, suggesting that the N-terminal domain is quite independent of the middle domain with the two domains joined by a flexible linker (Fig. 4.8b,c and 4.9). Interestingly, the closest structural homologues of the N-terminal domain of RGM_{CND} are the human MTCP1 oncogene (Fig. 4.8d) (PDB ID: 2HP8[305]), the MIT domain of VPS4-like ATPases (Fig. 4.8e) (PDB ID: 2V6Y[306]) and the complement inhibitor EHP from *S. aureus* (Fig. 4.8f) (PDB ID: 2NOJ[307]), despite very low sequence identities of 5 % or lower. The position of disulphide bridge II in RGM_{CND} is conserved in MTCP1 and a structural superposition of all four proteins shows that the fold is well conserved (Fig. 4.8g). These proteins span a wide range of functions suggesting that this triple helix bundle forms a versatile interacting unit.

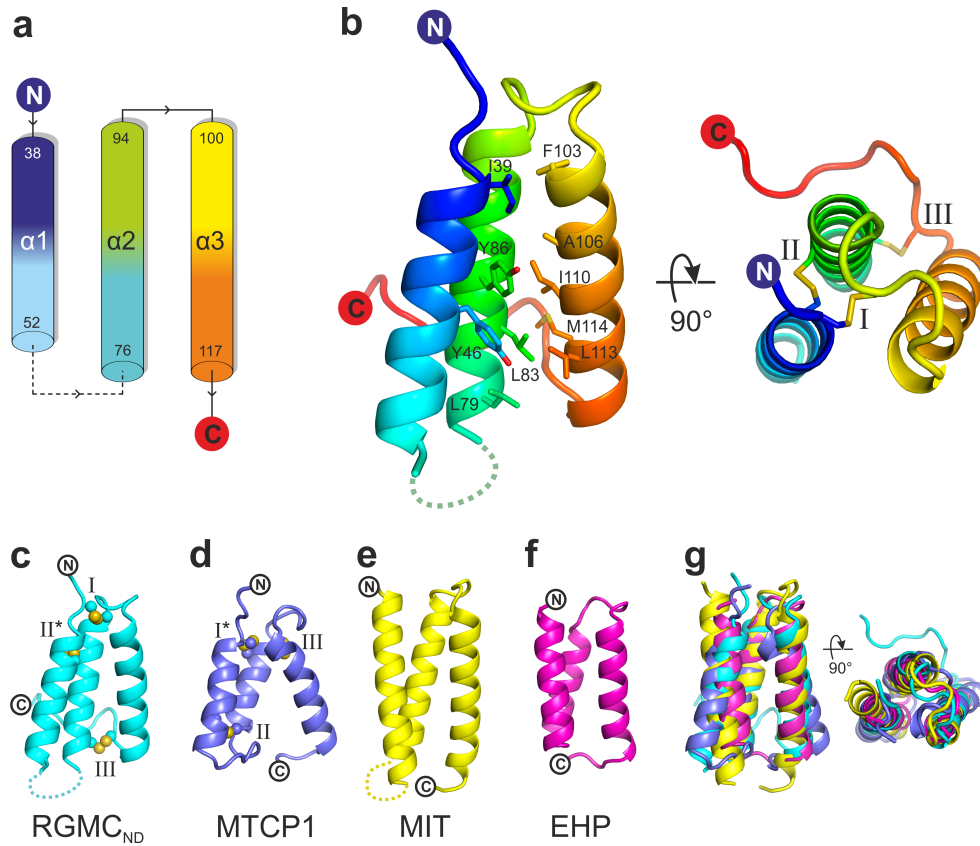


Figure 4.8: RGMC_{ND} structure and comparison. Reprinted by permission from Macmillan Publishers Ltd: [Nature Structural and Molecular Biology](#), [320] copyright (2015). **(a)** Topology diagram of RGMC_{ND} (adapted from PDB-SUM (<http://www.ebi.ac.uk/pdbsum/>)) in rainbow colouring (N-terminus: blue; C-terminus: red). N- and C-termini and residue numbers are shown. **(b)** Cartoon representation of the RGMC_{ND} structure. Colouring as in (a). RGM_{ND} forms a compact three helix bundle, stabilised by three disulphide bonds, supported by a hydrophobic core including residues from all three helices, depicted as sticks. Right panel rotated 90° around the x-axis compared to left panel. Disulphide bridges are numbered in Roman numerals. The disordered $\alpha 1$ - $\alpha 2$ loop is shown as a dotted line. **(c-g)** Structural comparison of RGMC_{ND} using the PDBfold server (<http://www.ebi.ac.uk/msd-srv/ssm/>). RGMC_{ND} ((c), cyan) shares the closest structural similarity to the human MTCP1 oncogene ((d), pdb 2HP8[305], slate; r.m.s.d. of 2.78 Å for 60 equivalent C α positions, sequence identity: 5 %), the MIT domain of VPS4-like ATPases ((e), pdb 2V6Y[306], magenta, r.m.s.d. of 2.47 Å for 60 equivalent C α positions, sequence identity: 3 %) and the complement inhibitor EHP from *S. aureus* ((f), pdb 2NOJ[307], yellow, r.m.s.d. of 2.34 Å for 54 equivalent C α positions, sequence identity: 6 %). The position of disulphide bridge II in RGM_{ND} is conserved in MTCP1 and highlighted by an asterisk. A structural superposition is shown in (g). The two views are related by a rotation of 90° around the x-axis.

RGMC_{ND} binds at the “helix-finger cavity” region of BMP2 and contacts both molecules of the disulphide-linked BMP2 dimer with a total buried surface area of 1690 Å² (Fig. 4.9a-c). BMP2 adopts a canonical structure with the cysteine knot domain and intermolecular disulphide clearly well ordered in the structure. The RGM-binding interface on BMP2 is conserved across all vertebrate BMP2, BMP4, BMP6 and BMP7 family members indicating the importance of this interaction (Fig. 4.9d). The RGMC_{ND}-BMP2 interface displays mixed electrostatics and is composed of six hydrogen bonds and 105 non-bonded contacts (Fig. 4.9e).

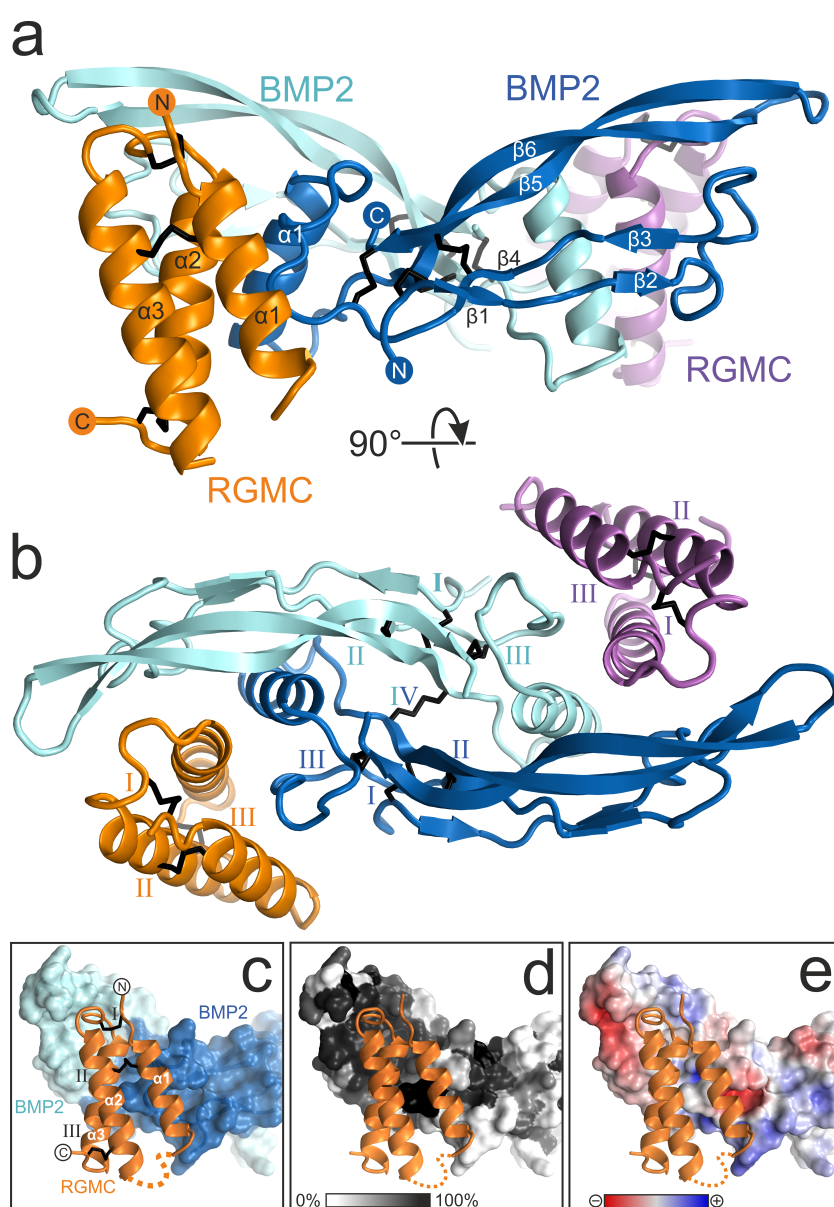


Figure 4.9: Structure of the BMP2-RGM complex. Reprinted by permission from Macmillan Publishers Ltd: [Nature Structural and Molecular Biology](#), [320] copyright (2015). (a) Cartoon representation of the BMP2-RGMC_{ND} complex. BMP2 is shown in blue and cyan, RGMC_{ND} in orange and violet. (b) View is rotated 90° around the x-axis compared to (a). Disulphide bonds (black sticks) are depicted with Roman numerals. (c-e) Surface representation of the BMP2 “finger” region showing the RGM binding site. RGMC is in cartoon representation. Disordered regions are shown as dotted lines. Orientation is as in (a). (c) The BMP dimer is colour-coded as in (b). (d) Amino acid residue conservation (from non-conserved, white, to conserved, black) is mapped onto the BMP2 dimer surface based on alignments containing sequences from all available vertebrate BMP2, BMP4, BMP6 and BMP7 family members. (e) Electrostatic potential from red (-8 kbT/ec) to blue (8 kbT/ec) is mapped onto the BMP2 dimer surface.

RGMA and RGM C both contain a conserved RGD motif (traditionally known to be important in integrin-fibronectin mediated adhesion[308, 59]). This motif, containing RGM C residues R98, G99 and D100, is located in a loop region between helices $\alpha 2$ and $\alpha 3$ and interestingly forms a major interaction site with BMP2 (Fig. 4.10a,b; residues highlighted by asterisks). Specifically, RGM C residues R98 and G99 provide several hydrogen bonds and non-bonded contacts that stabilise the position of the BMP2 finger loop (Fig. 4.10b). In particular, residue G99^{RGM C} is surrounded by many bulky residues contributed by BMP2, including Y373^{BMP2} and Y385^{BMP2}, which implies that the small glycine side-chain is important to allow for the tight packing observed at this interface. In addition, D100 of the RGD motif forms a hydrogen bond with A102 in RGM C thereby fixing the orientation of the RGM C $\alpha 2$ - $\alpha 3$ loop with respect to the $\alpha 3$ helix, which is important for the RGM C_{ND}-BMP2 interaction. In addition, residue R95^{RGM C} also forms a hydrogen bond with E376^{BMP2} at the very tip of the BMP finger, thereby stabilising the conformation of the BMP2 loop. R95^{RGM C} also forms a hydrogen bond through its carbonyl carbon with the side-chain of H104^{RGM C}, which positions the histidine to allow the formation of a π -stacking interaction with W313^{BMP2} (Fig. 4.10a,c). A T-shaped, orthogonal π -stacking between BMP2 residues W313^{BMP2} and W310^{BMP2} further stabilises this arrangement. The complex interface is therefore composed primarily of several key hydrogen bonds and a π -stacking network. Intriguingly, the JHH-linked residues G99 (of the proposed RGD motif) and L101 are located directly in the centre of the BMP-binding interface (Fig. 4.10b).

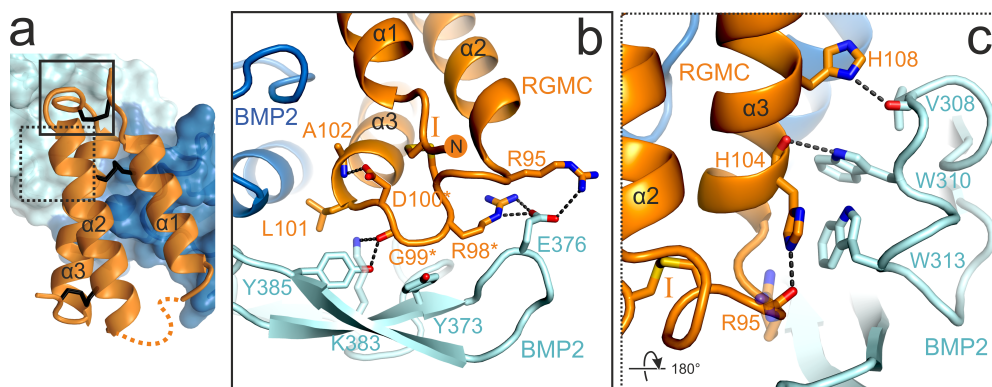


Figure 4.10: The BMP2-RGM complex interface. Reprinted by permission from Macmillan Publishers Ltd: *Nature Structural and Molecular Biology*, [320] copyright (2015). (a) Surface representation of the BMP2 dimer (blue and cyan) showing the RGMC (orange) binding site. Two regions containing key RGMC-BMP2 interactions are highlighted by boxes. (b-c) Close-up view, in cartoon representation, of the BMP2-RGMC interface. Colour-coding is as in (a). Selected interface residues are shown in stick representation. Hydrogen bonds are depicted as dotted lines. (b) The “RGD” motif is highlighted by asterisks. (c) The π -stacking interaction between RGMC-H104 and BMP2-W313 is shown.

4.3.3 The BMP2-binding mode is conserved across the RGM family

The crystal structures of BMP2 in complex with the N-terminal domains of human RGMA and RGMB were also solved, to 3.2 Å and 2.7 Å resolution respectively (Fig. 4.11, Table 1). A structural superposition of the three RGM_{ND}-BMP complex structures revealed that the overall fold of the RGM N-terminal domain, the RGM_{ND}-BMP2 complex architecture and the interface interactions are highly conserved across the RGM family members demonstrating a common mode for BMP-RGM binding (Fig. 4.11). The α 1- α 2 loop that was disordered in the RGM_{CND}-BMP2 complex was ordered in the RGMB_{ND}-BMP2 complex structure. This is due in part to this loop being shorter in RGMB compared to the other family members, and also to stabilising crystal contacts. This loop does not contact the BMP2 molecules and extends away from the RGM triple helix bundle, and is likely to be highly flexible in solution. Interestingly, the residues

corresponding to R98, G99, L101 and H104 in RGMC, that form key interactions with BMP2 in the complex, are conserved across the RGM family (Fig. 4.11b-e). RGMB is the only family member that does not contain the conserved RGD motif; instead the aspartic acid is substituted for an asparagine. However, in the complex structure N102^{RGMB} is oriented in a similar manner to D100^{RGMC} and forms a similar hydrogen-bonding pattern, showing that the important interactions are conserved (Fig. 4.11c). Other minor differences in residues around the periphery of the binding interface could account for the small differences in binding affinities observed for the RGM family members and BMP2[302] (Fig. 4.13).

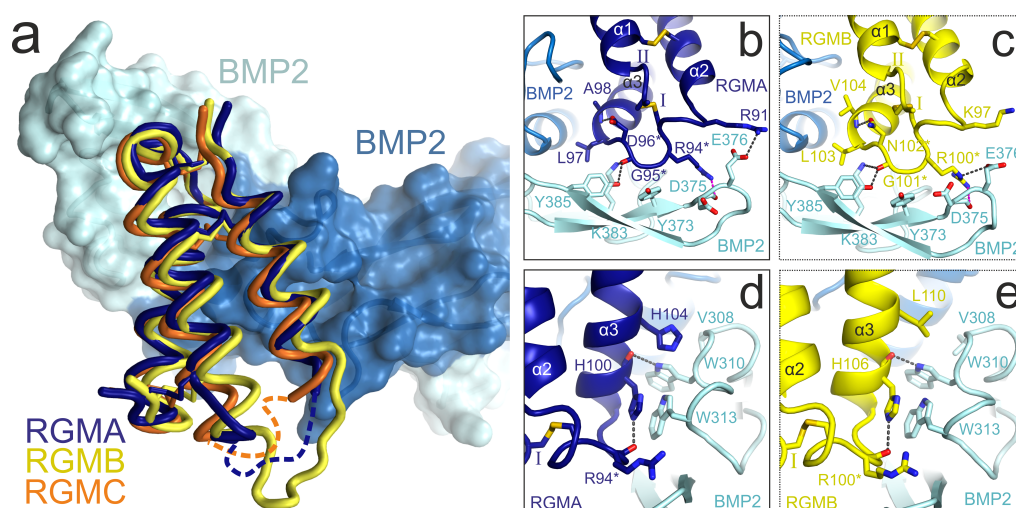


Figure 4.11: The mode of RGM-BMP2 interactions is conserved in RGMA, RGMB and RGMC. Reprinted by permission from Macmillan Publishers Ltd: *Nature Structural and Molecular Biology*, [320] copyright (2015). (a) Superposition of BMP2-RGMA_{ND}, -RGMB_{ND} and -RGMC_{ND} complex structures with the BMP2 dimer as reference. BMP2 dimer is shown as solvent accessible surface. RGMA (blue), RGMB (yellow) and RGMC (orange) are depicted as ribbons. Disordered loops are shown as dashed lines. (b-e) Close-up views of the RGM-BMP2 interfaces. Labelled residues are shown as sticks. The “RGD” motif is indicated by asterisks in (b-c). Note that the aspartate in the “RGD” motif is replaced with an asparagine in RGMB (c). Both the RGM “RGD” region and the π -stacking interactions (mediated via RGM helix α 3 and the two BMP2 residues W310 and W313) are conserved in all RGMs. Disulphide bonds are labelled with Roman numerals. Dashed lines indicate hydrogen bonds.

4.3.4 Validation of the RGM_{ND}-BMP2 interface

To validate the RGM_{CND}-BMP2 binding interface observed in the complex structures site-directed mutagenesis and SPR equilibrium binding experiments were performed (Fig. 4.13). The N-terminal domain of RGMs was confirmed to be both necessary and sufficient for the RGM-BMP2 interaction across all of the RGM family members. Both the full length RGM ectodomains (eRGMA, eRGMB, eRGM_C) and the RGM N-terminal domains (RGM_{AND}, RGM_{BND}, RGM_{CND}) bound to BMP2 with similar nanomolar affinities (Fig. 4.13a-f), and the tightest interaction was observed between BMP2 and RGM_{BND} with a K_d of 88 nM (Fig. 4.13e). This demonstrated that the RGM N-terminal domain is sufficient for interaction with BMP2. In contrast, the binding affinity between BMP2 and an RGMB construct lacking the N-terminal domain (RGMB_{ΔND}) was found to be greatly reduced with a K_d of 12 μ M compared to that of 415 nM for the eRGMB-BMP2 interaction, suggesting that the N-terminal domain of RGMs is necessary for tight interaction with BMP2 (Fig. 4.13g). These SPR analyses revealed a distinct level of non-specific interactions between eRGMs and BMP2, an effect that was markedly decreased for the constructs comprising just the N-terminal domain. Furthermore, the binding curve for the interaction between RGMB_{ΔND} and BMP2 was almost linear, resembling that of a non-specific interaction. This suggests that the C-terminal domain is responsible for the non-specific contribution observed in the eRGM SPR binding curves. To investigate this, the binding data for the eRGM_C-BMP2 interaction was fitted with both a 1:1 specific binding model and a 1:1 “total” binding model that also incorporates an element of non-specific binding (Fig. 4.12a). The fit for the 1:1 “total” binding model ($R^2 = 0.9985$) was better than that for the specific binding model ($R^2 = 0.9856$), which justifies the use of this model for the rest of the RGM SPR data. In addition, SPR kinetic experiments were also performed for the eRGMB-BMP2 and RGM_{BND}-BMP2 interactions. The kinetic binding

data for the RGM_{B_{ND}}-BMP2 interaction could be fitted with a 1:1 Langmuir binding model yielding a K_d of 92 nM (Res. sd = 2.46), similar to the K_d of 88 nM determined from the equilibrium binding analysis (Fig. 4.12b). However, the kinetic binding data for the eRGM-BMP2 interaction could not be fitted with a 1:1 Langmuir binding model (Res. sd = 21.32), seemingly due to the non-specific component of this interaction (Fig. 4.12c). This rationalises the use of equilibrium binding experiments to analyse the RGM-BMP2 interactions and further supports the use of the 1:1 “total” binding model for determining affinity parameters.

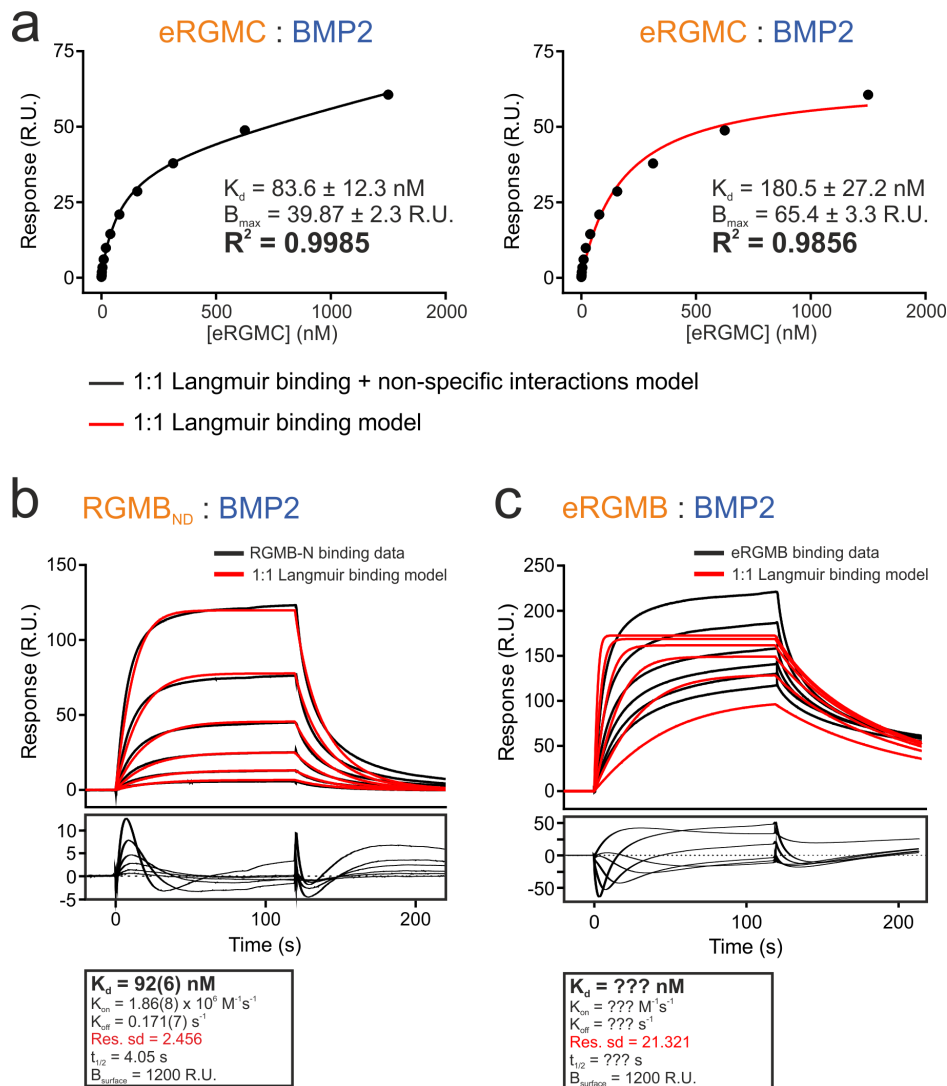


Figure 4.12: The eRGM-BMP2 interaction includes a non-specific component. (a) SPR equilibrium binding analysis of the eRGMC-BMP2 interaction. Different concentrations of eRGMC were injected over surfaces coupled with BMP2 (150 R.U.). The data were fitted to either a 1:1 Langmuir binding model + non-specific interactions (left panel, black curve) or a 1:1 Langmuir binding model (right panel, red curve). The binding parameters (K_d and B_{max}) derived from each model are shown. The R^2 value indicates the fit between model and data. (b-c) SPR kinetic analysis of RGMB_{ND} (b) and eRGMB (c) binding to BMP2. Graphs show a plot of binding response (response units (R.U.)) against time (s) for six different concentrations of analyte (2.4 nM, 4.9 nM, 9.8 nM, 19.5 nM, 39.0 nM, 78.1 nM). A 1:1 Langmuir binding model (red lines) is fitted to the binding curves (black lines). A plot of the residuals between each binding curve and the fitted model are displayed under each graph. The fitted parameters are displayed. The standard deviation of the residual (Res. sd) gives an indication of the goodness of fit between the data and the model. For eRGMB-BMP2 binding, the fit of the model was so poor (Res. sd = 21.321) that values for the parameters were not determinable.

Next, site-directed mutagenesis of residues that were identified as important for the RGM_{ND}-BMP2 binding interaction was performed. Residue H104^{RGMC} was targeted for mutagenesis studies as this formed a crucial part of the π -stacking network directly at the BMP2-binding interface. In an SPR equilibrium binding experiment, an RGMB H106R (equivalent to H104 in RGMC) mutation was found to reduce the binding affinity for BMP2 by approximately 100 fold, from 88 nM for RGMB_{ND} to 8.5 μ M for RGMB_{ND}-H106R (Fig. 4.13h). This value is comparable to the affinity of the interaction between RGMB _{Δ ND} and BMP2 (K_d = 12 μ M), suggesting that this mutation abolishes RGMB_{ND}-BMP2 binding, thereby validating the observed binding mode. The H106R mutation introduces a bulky, charged side-chain at the interface and so to dissect the contribution of the H104^{RGMC}-W313^{BMP2} π -stacking to the overall binding affinity, an RGMC H104A mutant was tested. This mutation disrupts the π -stacking without sterically hindering complex formation and the binding affinity was reduced from 124 nM for RGM_{ND} to 280 nM for RGM_{ND}-H104A, suggesting that the π -stacking network contributes two-fold to the RGM-BMP2 binding affinity (Fig. 4.13i).

SPR equilibrium binding experiments were also used to analyse the BMP-binding affinity of the JHH-linked RGM C mutations G99R and L101P that are located at the RGM C-BMP2 interface. These mutations severely disrupted RGM C-BMP2 binding; the G99R mutation reduced the affinity by roughly ten-fold, from 120 nM for RGM C_{ND} to 910 nM, and the L101P mutation decreased the binding affinity by approximately 100-fold, to 1.5 μ M (Fig. 4.13j,k). An analysis of the structure rationalises these effects: the mutation G99R would cause side-chain clashes at the centre of the interface and the mutation L101P, due to the restricted ϕ and ψ bond angles available to proline, would most likely disrupt both the orientation of the α 2- α 3 BMP-binding loop and the folding of the α 3 helix. These binding data not only validate the interface observed in the RGM_{ND}-BMP2 structures, but also suggest that disruption of the RGM C-BMP6 interaction is the molecular mechanism for these mutations in JHH disease pathology.

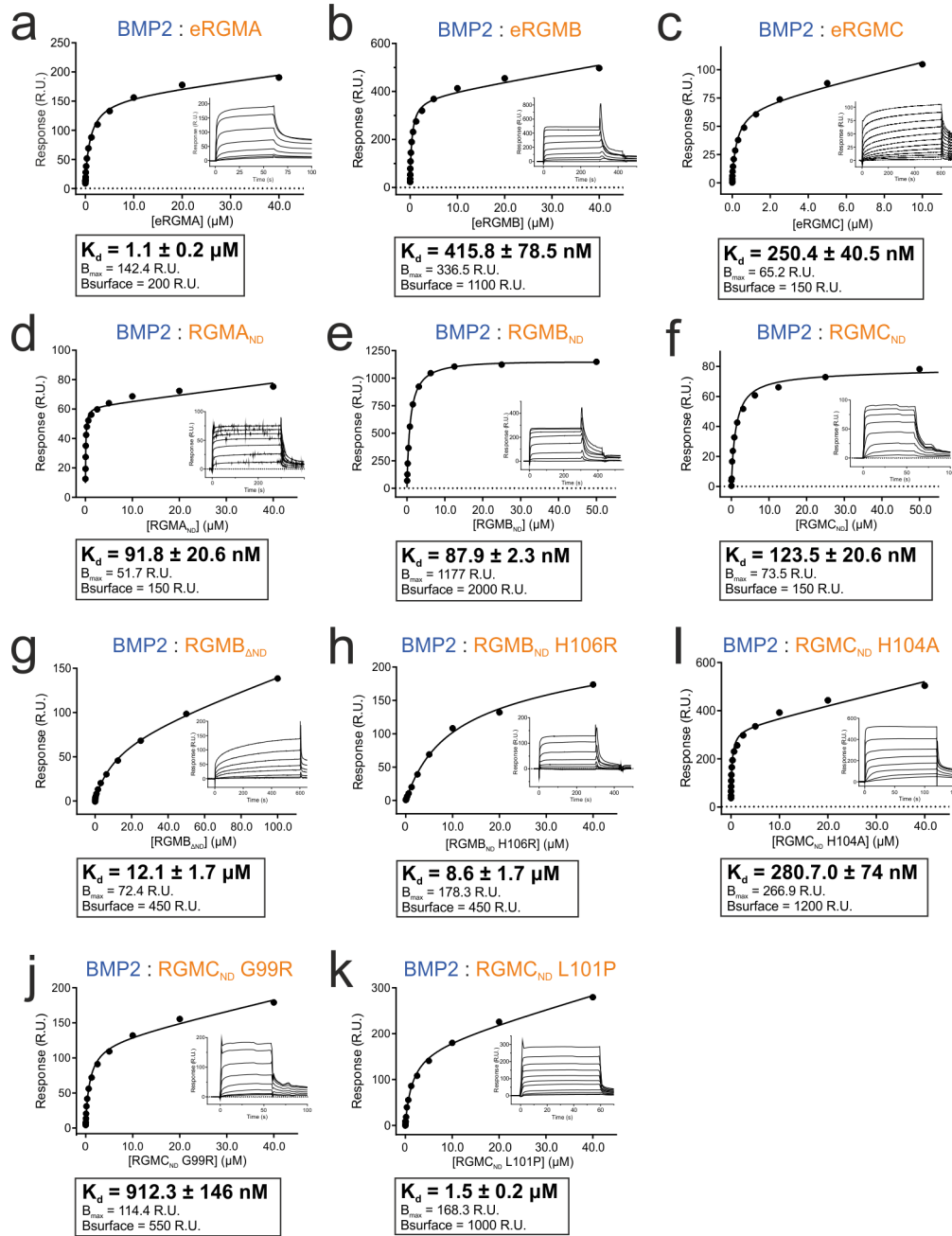


Figure 4.13: RGM-BMP2 SPR equilibrium binding experiments. Reprinted by permission from Macmillan Publishers Ltd: [Nature Structural and Molecular Biology](#), [320] copyright (2015). Binding of different constructs and mutants of RGMA, RGMB and RGMC, respectively, to BMP2. Graphs show a plot of the equilibrium binding response (response units (R.U.)) against concentration of the analytes used (μM). Binding constants (K_d) are given as the mean $\pm$ the standard error of the mean ($n=2$ technical repeats) Fits are shown as lines. Sensorgrams are shown as plots of response (R.U.) against time (s). Binding parameters K_d , B_{max} and surface response units (B_{surface}) are indicated below the graphs.

4.3.5 The structure of the eBMPR1A-BMP2 complex reveals new features

In addition to the RGM-BMP2 complex structures, the structure of eBMPR1A in complex with BMP2 was also solved (Fig. 4.6, 4.14 and Table 1). A superposition of this structure with the two previously solved eBMPR1A-BMP2 structures (PDB 1ES7[241] Fig. 4.14c, PDB 2GOO[298] Fig. 4.14d) shows that the complexes are very similar (r.m.s.d. = 0.254). However, in contrast to these structures, where eBMPR1A was refolded from bacterial inclusion bodies, in this work eBMPR1A was expressed in mammalian cells. Therefore, for the first time, the glycosylation sites on eBMPR1A are visible in the structure (Fig. 4.14b). One O-linked glycosylation site at T137^{BMPR1A} and one N-linked glycosylation site at N37^{BMPR1A} were clearly resolved (Fig. 4.14f). At T137^{BMPR1A} one N-acetylgalactosamine residue linked via a β -1,3-glycosidic bond to a galactose was built, in accordance with the standard scheme of mammalian O-linked glycosylation; and at N37^{BMPR1A} a single N-acetylglucosamine was built, as expected for the first residue in canonical mammalian N-linked glycosylation (Fig. 4.14f).

There were four copies of the eBMPR1A-BMP2 complex in the asymmetric unit and in one of those copies an α -helix could be built at the N-terminus of one BMP2 monomer, comprising residues A284-L292 (Fig. 4.14b,e). These N-terminal residues were not visible in the previously solved eBMPR1A-BMP2 complex structures (PDB 1ES7[241], PDB 2GOO[298]), however, they were resolved in structures of the closely related ligand TGF β , where they also take the form of an α -helix[285]. This novel BMP2 N-terminal helix forms several hydrogen bonds and non-bonded contacts with eBMPR1A, most notably the hydrogen bonds between R291^{BMP2} and E87^{BMPR1A}, although whether this helix is important for BMP2 binding to BMPR1A remains to be seen (Fig. 4.14e).

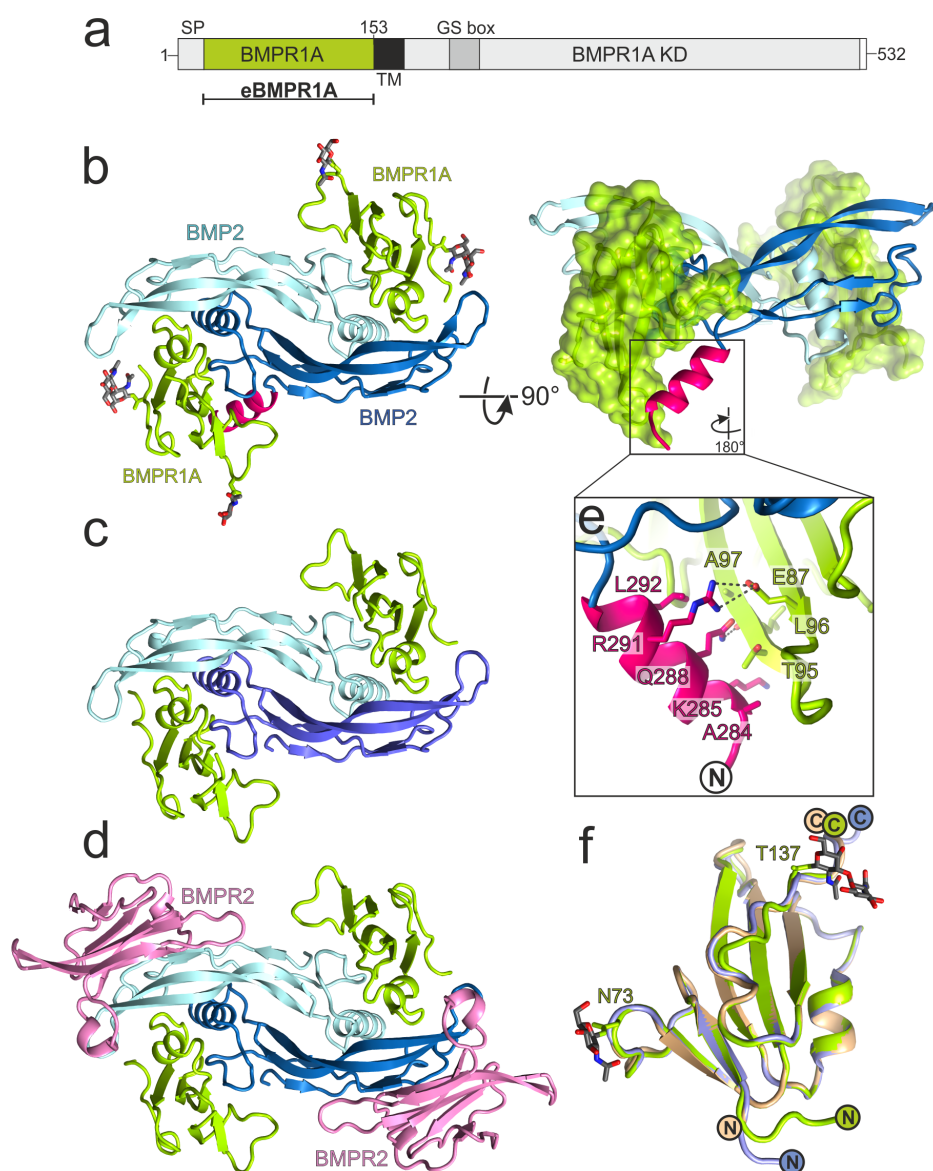


Figure 4.14: The structure of the eBMPR1A-BMP2 complex. (a) Schematic of human BMPR1A, the eBMPR1A construct is indicated. GS box = site of activation, TM = transmembrane helix, KD = kinase domain. (b) Cartoon representation of the eBMPR1A-BMP2 complex. BMP2 = blue/cyan, eBMPR1A = green, the novel BMP2 N-terminal helix = pink. Sugars shown as sticks and coloured by element. Right panel rotated 90° about x-axis and eBMPR1A is shown as surface. (c-d) Cartoon representation of the published eBMPR1A-BMP2 (PDB ID: 1ES7[241]) (c) and eBMPR1A-BMP2-eACVR2A (PDB ID: 2GOO[298]) (d) structures, eACVR2A = pink. (e) Close-up view of the BMP2 N-terminal helix interactions with eBMPR1A. Interface residues shown as sticks, hydrogen bonds shown as dotted lines. Orientation rotated 180° about the y-axis from (b), right panel. (f) Superposition of eBMPR1A structures from this work (green), 1ES7 (slate) and 2GOO (peach) (r.m.s.d. = 0.283). Glycans shown as sticks, N- and C-termini labelled.

4.3.6 RGM competes with BMPR1A for BMP2-binding

To probe the role for RGMs in BMP signalling, SPR equilibrium binding experiments were used to test for RGM binding to the BMP receptors (Fig. 4.15). The binding affinities for the BMP receptors to the BMP2 ligand was found to be in the expected range, eBMPR1A bound to BMP2 with a K_d of 280 ± 10 nM (Fig. 4.15a), a value that is in agreement with previous studies[294], and the binding affinities for eBMPR2 and eACVR2A were found to be around 10 fold lower ($K_d = 980$ nM and $K_d = 35 \mu\text{M}$, respectively), as expected (Fig. 4.15b,c). In contrast, using our experimental setup, no binding was detected between eRGMB and any of the BMP receptor ectodomains (Fig. 4.15e-g); this result differs from that in previously reported pull-down experiments which showed an interaction between RGMs and full-length BMP receptors[89]. These binding studies suggest that the mechanism of activation of BMP signalling by RGMs is not through direct contact with the BMP receptors.

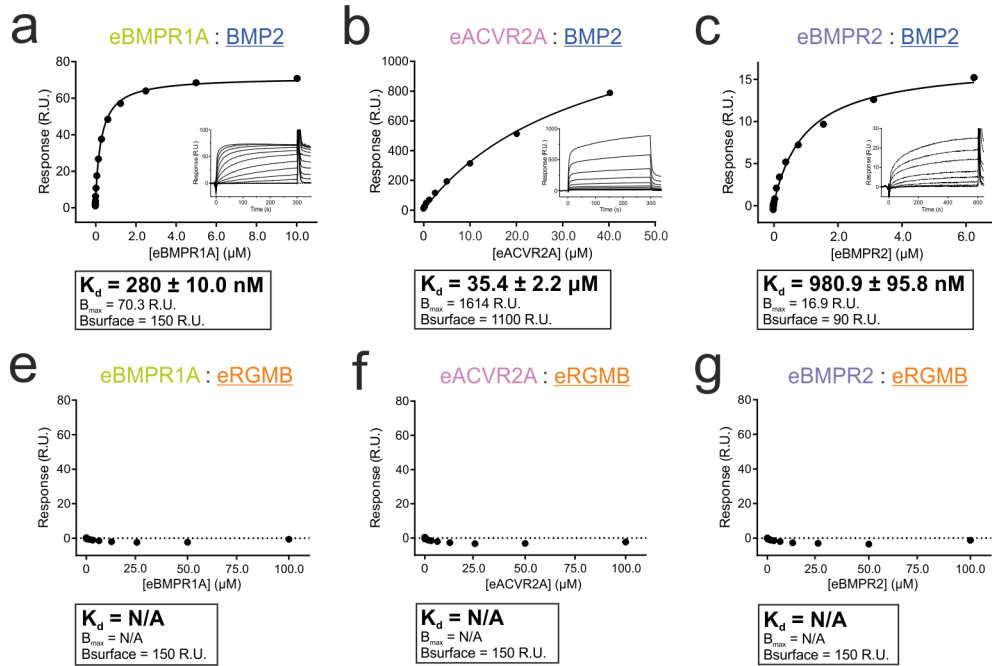


Figure 4.15: BMP-receptors SPR equilibrium binding experiments. Binding of the BMP receptors eBMPR1A, eACVR2A and eBMPR2 to BMP2 (**a-c**) or eRGMB (**e-g**). Each graph is labelled with the analyte and ligand. The binding partner immobilised on the chip is underlined. Graphs show a plot of the equilibrium binding response (response units (R.U.)) against concentration of the analytes used (μM). All experiments were performed in duplicate. Binding constants (K_d) are given as the mean with the error range the standard error of the mean ($n=2$ technical repeats) Fits are shown as lines. Sensorgrams are shown as plots of response (R.U.) against time (s). Binding parameters K_d , B_{max} and surface response units (B_{surface}) are indicated below the graphs.

A superposition of the RGM_{ND} -BMP2 and the published eBMPR1A-BMP2-ACVR2A (PDB ID: 2GOO[298]) complexes demonstrated that RGM_{ND} unexpectedly shares an overlapping BMP2-binding interface with eBMPR1A (Fig. 4.16a-c). Despite the very different folds of the RGM_{ND} and eBMPR1A ligands, the BMP2-binding sites are remarkably similar. The BMPR1A $\alpha 1$ helix, which was identified as the crucial interaction determinant in the eBMPR1A-BMP2 complex[241], is mimicked by the RGM $\alpha 3$ helix (Fig. 4.16a). A close-up analysis of the interfaces shows that both BMPR1A and RGM form π -stacking interactions with BMP2 tryptophan residues, with F108^{BMPR1A} interacting with W310^{BMP2} and H104^{RGM} with W313^{BMP2} (Fig. 4.16d). This helix $\alpha 1$ -hydrophobic pocket motif was identified as crucial for type I receptor binding to BMPs and a large hydrophobic residue at the site of F108^{BMPR1A} features in all type I receptors (except ALK1)[241]. It appears that RGMs have employed a similar motif for BMP-binding. In contrast, the BMP2-binding site for the BMP type II receptors (eBMPR2, eACVR2A) does not overlap with that for RGMs (Fig. 4.16a).

Given the similar affinities of the RGMB-BMP2 and eBMPR1A-BMP2 interactions and their overlapping binding interfaces on the BMP2 ligand, an SPR competition experiment was carried out to investigate whether RGMB and BMPR1A compete for BMP2 binding (Fig. 4.16e). In this experiment a constant concentration of eRGMB together with an increasing concentration of eBMPR1A was

injected over BMP2 bound to the SPR chip surface. A decrease in the response units indicated binding of eBMPR1A to BMP2 rather than eRGMB, as the receptor is approximately half the molecular weight of eRGMB (24 kDa vs. 43 kDa). Therefore, an IC_{50} value could be calculated for this competition-binding curve ($R^2 = 0.87$), which indicated that approximately a 4 μ M concentration of eBMPR1A is required to outcompete a 2.5 μ M concentration of eRGMB for half of the BMP2 binding sites (Fig. 4.16e). This suggests that a 1.7 times molar excess of eBMPR1A was required to effectively displace eRGMB in solution, suggesting that RGM and BMPR1A are likely to compete for BMP binding *in vivo*.

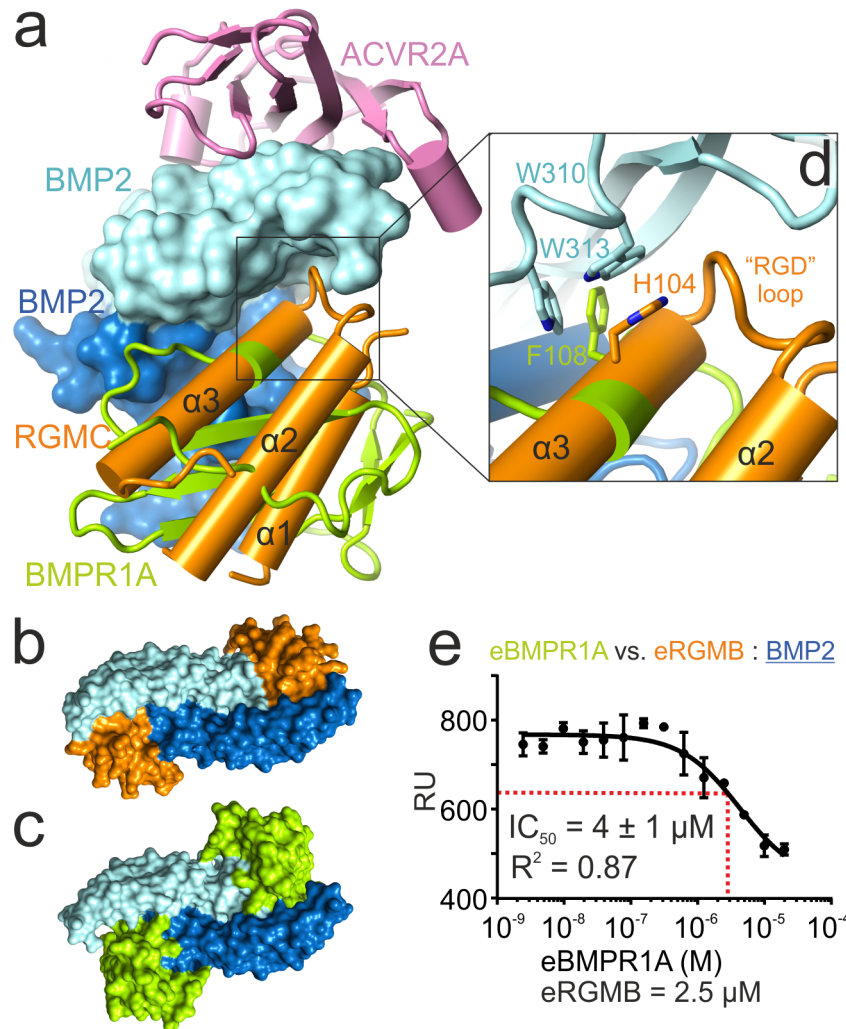


Figure 4.16: RGM and BMPR1A compete for binding to BMP2. Reprinted by permission from Macmillan Publishers Ltd: [Nature Structural and Molecular Biology](#), [320] copyright (2015). (a) BMP2-RGMC_{ND}, and BMP2-eBMPR1A-eACVR2A (pdb code 2GOO[298]) complexes are superimposed. BMP2 surface (light and dark blue) and RGMC_{ND} (orange), eBMPR1A (green) and eACVR2A (pink) ribbons are shown. (b-c) Surface representation of the RGMC_{ND}-BMP2 (b) and eBMPR1A-BMP2 (c) complexes. (d) Close-up of the BMP2-binding region of RGM and BMPR1A; the residues involved in π -stacking interactions are labelled. (e) SPR competition binding experiment between eBMPR1A and eRGM to immobilised BMP2. An IC₅₀ value of 4.0 ± 1.0 μM was calculated. R²=goodness of fit. Error bars, s.e.m. (n=3 technical replicates).

4.3.7 The RGM-BMP2 interaction is pH-dependent

Despite the similarity of the BMPR1A and RGM BMP2-binding epitopes there are crucial differences between the interfaces, for example the important π -stacking networks involve a histidine in RGMB (H106^{RGMB}) but a phenylalanine in BMPR1A (F108^{BMPR1A}) (Fig. 4.16d). Given that the side chain of histidine has a pKa value of approximately 6.5 and histidine residues are well known for being readily protonated within biological systems, the question was posed, whether the RGMB-BMP2 interface might be pH-sensitive. To test this hypothesis, multi-angle light scattering (MALS) measurements of both the RGMB_{ND}-BMP2 and eBMPR1A-BMP2 complexes were performed at different pH values (Fig. 4.17a,b). For RGMB_{ND}-BMP2, a major species was observed at neutral pH with an experimentally determined molecular weight of 46.2 ± 2.7 kDa, which corresponds to the 2:2 RGMB_{ND}-BMP2 complex with a calculated molecular weight of 47 kDa (Fig. 4.17a). A second minor peak with an experimentally determined molecular weight of 22.4 ± 0.9 kDa was also observed that was attributed to excess unbound BMP2 (calculated weight 26 kDa). However, at pH 6.5 and pH 5.5 the peak corresponding to the RGMB_{ND}-BMP2 complex was no longer detected, and instead two peaks with experimentally determined molecular weights of 20.1 ± 1.4 kDa and 13.4 ± 0.4 kDa at pH 6.5 and 24.0 ± 1.2 kDa and 16.2 ± 0.3 kDa at pH 5.5 were observed, which most likely correspond to the free monomeric proteins (calculated weights BMP2 26 kDa, RGMB_{ND} 11 kDa) (Fig. 4.17a). In contrast, for the eBMPR1A-BMP2 complex, a major species with an experimentally determined weight of 53.0 ± 0.1 kDa at pH 7.5, 53.2 ± 0.3 kDa at pH 6.5 and 54.2 ± 0.5 kDa at pH 5.5 was observed, which corresponds to the 2:2 eBMPR1A-BMP2 complex (calculated weight 53 kDa) (Fig. 4.17b). A second peak was also present across the pH range with an experimental weight of 15.0 ± 0.2 kDa at pH 7.5, 18.8 ± 0.1 kDa at pH 6.5 and 19.7 ± 0.4 kDa at pH 5.5, which was assumed to be excess eBMPR1A (calculated

weight 14 kDa) (Fig. 4.17b). Taken together, these data show that, unlike the 2:2 eBMPR1A-BMP2 complex, the RGMB_{ND}-BMP2 interaction is pH sensitive, with the complex dissociating at pH 6.5 and below.

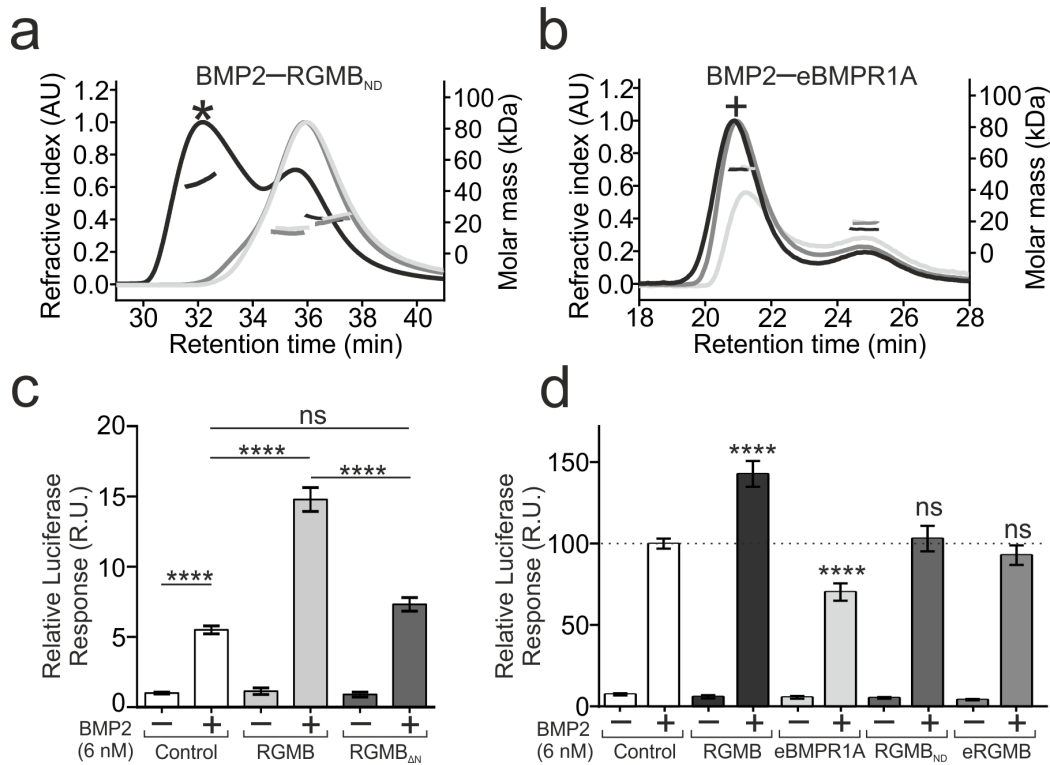


Figure 4.17: The eRGMB- and eBMPR1A-BMP2 complexes behave differently in functional assays. Reprinted by permission from Macmillan Publishers Ltd: [Nature Structural and Molecular Biology](#), [320] copyright (2015). **(a-b)** MALS analysis of the BMP2-RGMB_{ND} **(a)** and the BMP2-eBMPR1A **(b)** complexes at pH 7.5 (black), pH 6.5 (dark grey) and pH 5.5 (light grey). Absorbance curves (at 280 nm) are plotted and lines indicate the molecular weights (kDa) of the peaks. The BMP2-RGMB_{ND} complex (46.2 ± 2.7 kDa) is marked by an asterisk and the BMP2-eBMPR1a complex ($53.0 \pm 0.1/53.2 \pm 0.3/54.2 \pm 0.5$ kDa) is marked by a plus sign. **(c)** BRE-Luc SMAD-mediated BMP2 response assay with co-transfected GPI-anchored RGMB constructs; three independent experiments. Error bars, s.e.m. ($n = 40$ except RGMB_{ΔND}-BMP2 where $n = 20$ cell cultures), **** $P < 0.0001$, ns = 0.0509 by one-way ANOVA with a Tukey multiple comparison test. **(d)** BRE-Luc SMAD-mediated BMP2 response assay with co-transfected eBMPR1A or RGMB ectodomain constructs; two independent experiments. Error bars, s.e.m. ($n = 37, 72, 8, 23, 8, 22, 8, 24, 24, 22$ left to right), **** $P < 0.0001$, ns = 0.992 (RGMB_{ND}), 0.861 (eRGMB) by one-way ANOVA with Tukey multiple comparison test.

4.3.8 RGMB and BMPR1A have different effects on cellular BMP signalling

Given that RGMB and BMPR1A compete for BMP2 binding the question is: how then can RGM function as an activator of BMP signalling if it occludes the BMP type I receptor binding site? The RGMB_{ND}-BMP2 complex is pH-sensitive whereas the eBMPR1A-BMP2 complex is not, suggesting that the two complexes may behave differently *in vivo*. To test the effects of both proteins on BMP signalling in cells a BMP-responsive luciferase reporter (BRE-Luc) assay was conducted using LLC-PK1 cells[267, 89]. Stimulation with 6 nM purified BMP2 increased BRE-luciferase activity by ~5-fold (n=40, $P<0.0001$) over the control (Fig. 4.17c). When the cells were transfected with full-length (GPI-anchored) RGMB, BRE-luciferase activity was further enhanced by ~3-fold (n=40, $P<0.0001$) (Fig. 4.17c), a level of activation that was comparable to published results[89]. In contrast, transfection with a GPI-anchored RGMB construct lacking the N-terminal domain (RGMB_{ΔND}) had no effect (n=40, $P=0.0509$) on the BMP-induced response (Fig. 4.17c). This confirmed that the RGM N-terminal domain is necessary for RGM-mediated activation of BMP signalling, in agreement with structural and SPR analysis.

Next, the effects of soluble proteins (lacking the membrane attachment sites) were investigated in the same luciferase reporter assay (Fig. 4.17d). In this experiment, stimulation with 6 nM purified BMP2 increased BRE-luciferase activity by ~10-fold (n=72, $P<0.0001$) over the control (n=37) (Fig. 4.17d). As expected, transfection with GPI-anchored RGMB further enhanced BMP signalling, here by ~1.4-fold (n=24, $P<0.0001$). When cells were transfected with a plasmid for the expression of soluble eBMPR1A, BRE-luciferase activity decreased to ~75% (n=22, $P<0.0001$), most likely because soluble eBMPR1A bound to available BMP2 ligands, preventing binding to endogenous type I receptors, and inhibit-

ing signalling. In contrast, transfection with plasmids for the expression of either soluble eRGMB or RGMB_{ND} had no effect on the BMP2-induced BRE-luciferase activity ($n=22$, $P=0.8606$; $n=24$, $P=0.992$ respectively) (Fig. 4.17d). This deviated from previous studies where a soluble RGMB-Fc fusion construct was shown to inhibit BMP signalling[89]. Therefore, to further validate this observation, similar luciferase assays were performed but using purified eBMPR1A or eRGMB proteins, which were added directly to either LLC-PK1 or C2C12 cells (Fig. 4.18). The EC50 value for BMP2 stimulation of the LLC-PK1 cells was ~ 14 nM (Fig. 4.18a), therefore, the luciferase assays were performed using three different BMP2 concentrations (6 nM, 10 nM and 25 nM) to span this value. A dilution series of eBMPR1A and eRGMB, ranging from 100 times the concentration of BMP2 down to approximately half the BMP2 concentration was used. This experimental design tested whether the observed effects were consistent across different cell lines and BMP2, eBMPR1A and eRGMB concentrations. Again, using these experimental setups, we observed inhibition of BMP signalling by eBMPR1A, whereas eRGMB did not inhibit BMP2 signalling in either cell type, even at concentrations of $2.5 \mu\text{M}$, which is three times the K_d for the eRGMB-BMP2 interaction (Fig. 4.18b-d). In summary, soluble eBMPR1A acts as a “ligand-trap”, competing with endogenous BMP type I receptors and inhibiting signalling, as expected. Surprisingly however, soluble eRGMB did not inhibit BMP signalling, despite eBMPR1A and eRGMB exhibiting similar binding affinities for the BMP2 ligand and competing for BMP2 binding in solution.

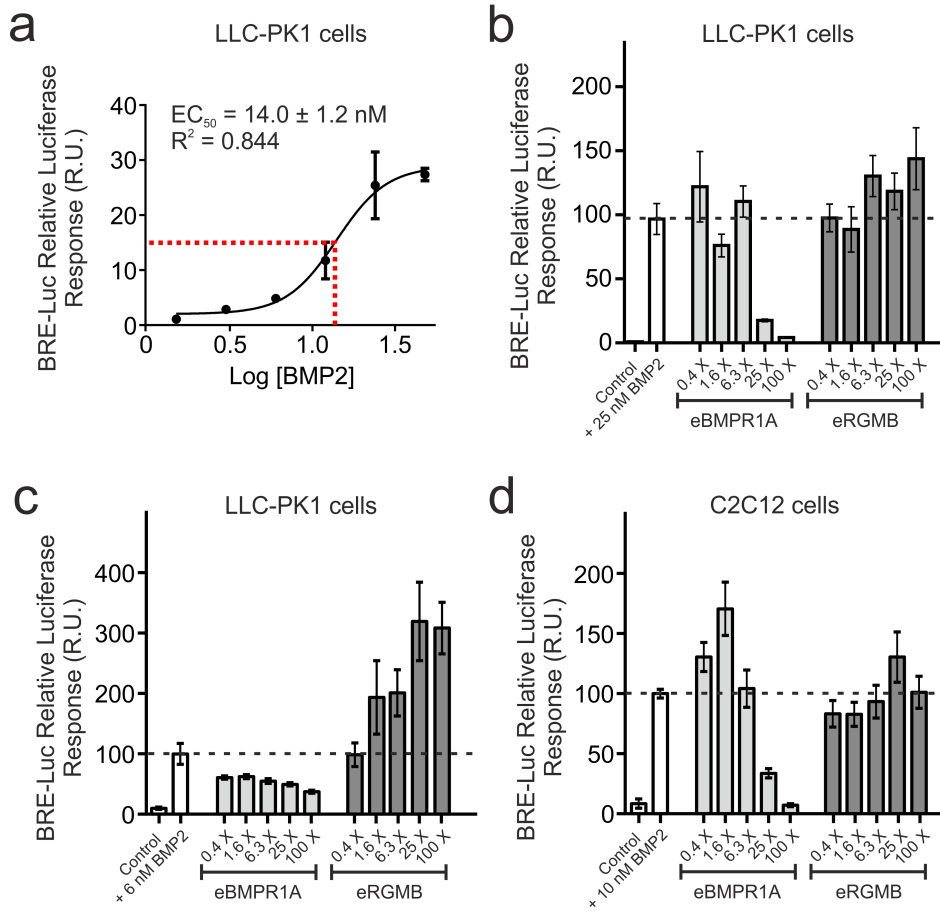


Figure 4.18: Analysis of the effects of eBMPR1A and eRGMB on BRE-Luc SMAD-mediated BMP2 response. (a) LLC-PK1 cells were stimulated with a range of BMP2 (1.5, 3, 6, 12, 24, 48 nM) concentrations. Graph shows a plot of Log₁₀[BMP2] against the relative luciferase response (R.U. = response units). The calculated (GraphPad Prism) EC₅₀ value is shown and is indicated by a dotted red line. The R² value indicates the fit between the model (black curve) and the data points. Data are from two independent experiments. Error bars are s.e.m., n = 6-8 in each group, where n = cell cultures. (b-d) LLC-PK1 cells (b, c) or C2C12 cells (d) were stimulated with either buffer control or different concentrations of BMP2: 25 nM (b), 6 nM BMP2 (c); or 10 nM BMP2 (d). BMP2 was pre-incubated with 0.4, 1.6, 6.3, 25 or 100 X molar excess of eBMPR1A or eRGMB. Average BRE-Luc relative response was calculated for each condition from two independent experiments. For (b): “Control” n = 4, “+25 nM BMP2” n = 12, all others n = 8. For (c): “Control” and “+ 6 nM BMP2” n = 33, all eBMPR1A n = 41, all eRGMB n = 32. For (d): “Control” n = 6, “+10 nM BMP2” n = 18, all others n = 12. Where n = cell cultures. Error bars are s.e.m., dotted line indicates average increase in relative luciferase response induced by the respective BMP2 concentration in each case.

4.3.9 Investigating internalisation of the BMP ligand

These structural and biophysical data implied a potential model for RGM-activation of BMP signalling that is linked to endocytosis of the BMP ligand. The internalisation of the BMP2 ligand has previously been observed in fixed HeLa cells using BMP2 conjugated to the fluorescent dye Dylight633, which was prepared using an NHS-activated Dylight633 amine-reactive dye protein labelling kit from Thermo Scientific[183]. Therefore, towards testing the effects of RGMs on the rate of BMP2 endocytosis in live cells this labelling procedure was replicated here. However, poor yields of fluorescently conjugated BMP2 were initially obtained following the manufacturer's guidelines and therefore the labelling protocol was optimised. The labelling reaction between the NHS ester moiety on the Dylight633 dye and the free primary amines on BMP2 is most efficient at pH 8.5; however, refolded BMP2 is stored in a pH 4.0 buffer and precipitation is observed at higher pH values. Therefore, BMP2 was mixed in a 1:1 ratio with RGMB_{ND}, which stabilised the BMP2, allowing the pH to be increased to pH 8.5 using the sodium borate labelling buffer supplied with the kit. Following this, the protein complex was labelled with Dylight633 according to the manufacturer's guidelines (see Materials and Methods 2.11.8) and then dialysed against a buffer containing 4 M Urea, 100 mM Tris-HCl, pH 6.0, 5 mM EDTA resulting in the dissociation of the BMP2-RGMB_{ND} complex and allowing the separation of the two Dylight633-conjugated proteins by SEC (Fig. 4.19a). Labelled fractions were identified by their blue colour and the identities of SEC peaks were determined by comparing the reducing and non-reducing SDS-PAGE analysis (Fig. 4.19b). The Dylight633-labelled BMP2 dimer was found to be active in a BRE-Luc luciferase reporter assay (Fig. 4.19c).

In order to monitor the rate of internalisation of BMP2-Dylight633, C2C12 cells were transfected with a plasmid encoding Rab5a-mCherry, an established marker

of the early endosome. The rate of co-localisation of BMP2-Dylight633 and Rab5a-mCherry would give an indication of the rate of endocytosis of the ligand. A time-course of images was recorded of live C2C12 cells transfected with Rab5a-mCherry after the addition of BMP2-Dylight633 (Fig. 4.19d). However, initial tests revealed several problems; first, it was difficult to resolve single Rab5a-mCherry labelled endosomes, making it difficult to quantitatively analyse co-localisation with the BMP2-Dylight633 (Fig. 4.19d). In addition, the emission spectra of the mCherry and Dylight633 fluorophores exhibit a great deal of overlap, resulting in bleed-through of the mCherry signal into the Dylight633 detector, which made it difficult to unambiguously determine internalised points as BMP2. Finally, the BMP2-Dylight633 conjugate seemed to be very sticky causing it to adhere to the coverslip on which the cells were growing (Fig. 4.19d). Even at very low concentrations of BMP2-Dylight633 a lot of background and some protein aggregation was observed, which again made it difficult to observe internalisation. Unfortunately, due to these limitations useable data quantifying internalisation of the fluorescently labelled BMP2 ligand could not be obtained.

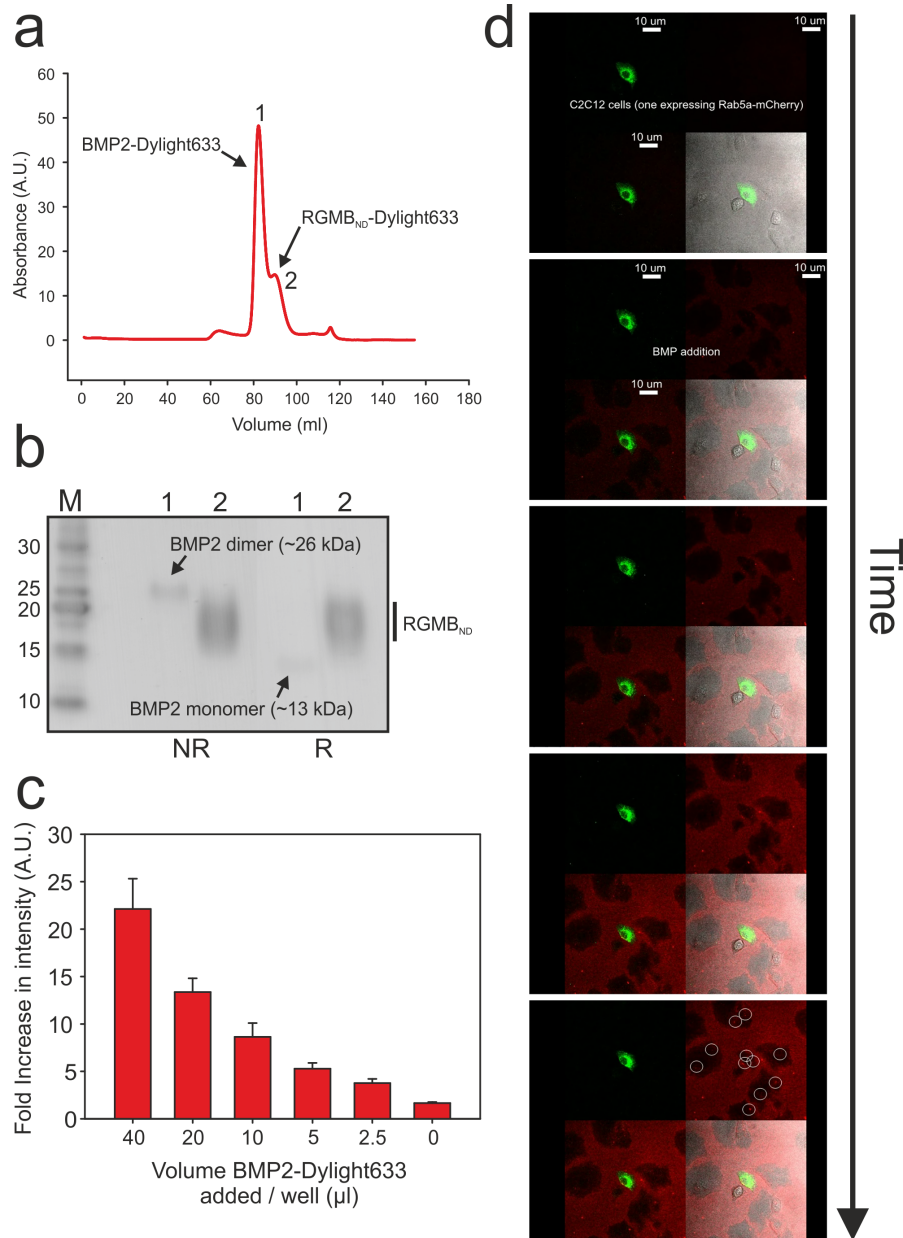


Figure 4.19: Preliminary investigation of BMP2-Dylight633 internalisation. (a) SEC separation of labelled BMP2-Dylight633 (peak 1) from RGMB_{ND}-Dylight633 (peak 2). (b) Reducing (R) and non-reducing (NR) SDS-PAGE analysis of SEC peaks 1 and 2. Peak 1 was identified as BMP2-Dylight633 due to visualisation of the monomer and dimer. (c) BRE-Luc BMP2 luciferase response assay over a range of added volumes (2.5-40 μl) of BMP2-Dylight633. Data from two independent experiments, n = 4 cell cultures for each volume, error bars are s.e.m.. (d) A time course (over 30 min) of fluorescence microscopy images of Rab5a-mCherry (green) transfected C2C12 cells following addition of BMP2-Dylight633 (red). Four different channels are shown: Rab5a-GFP (top, left), BMP2-Dylight633 (top, right), merge (bottom, left) and merge + transmitted light (bottom, right). Potential sites of BMP2-Dylight633 internalisation are indicated in the last image by white circles (top right panel).

4.3.10 Tracking the subcellular localisation of BMP-receptor complexes

Given the difficulties encountered in tracking the internalisation of the BMP2 ligand, an alternative approach analysing the subcellular location of the formation of BMP type I and type II receptor complexes was investigated. Endocytosis of the BMP ligand implies that active BMP signalling complexes will form primarily in the endosomal compartments, this hypothesis was tested by analysing the spatial localisation of receptor-receptor FRET in live cells after addition of BMP2 ligand. Towards this, several constructs of BMPR1A and BMPR2 labelled C-terminally, through a flexible linker, with mVenus or mTFP, a well-establish FRET pair, were generated. Two constructs of BMPR2 were cloned, a truncated form (BMPR2t) containing only the receptor kinase domain and a full-length form (BMPR2), which also included the 500 amino acid uncharacterised C-terminal region (Fig. 4.20a). The expression of the different BMPR1A and BMPR2 constructs was screened in C2C12 and HEK293T cells using both PEI and Lipofectamine[®] 2000 as transfection reagents (Fig. 4.20b). Eventually, BMPR1A-mVenus, BMPR2-mTFP and BMPR2t-mTFP were chosen for subsequent FRET-FLIM experiments in COS7 cells using Lipofectamine[®] 2000 as a transfection reagent due to slightly enhanced transfection efficiency.

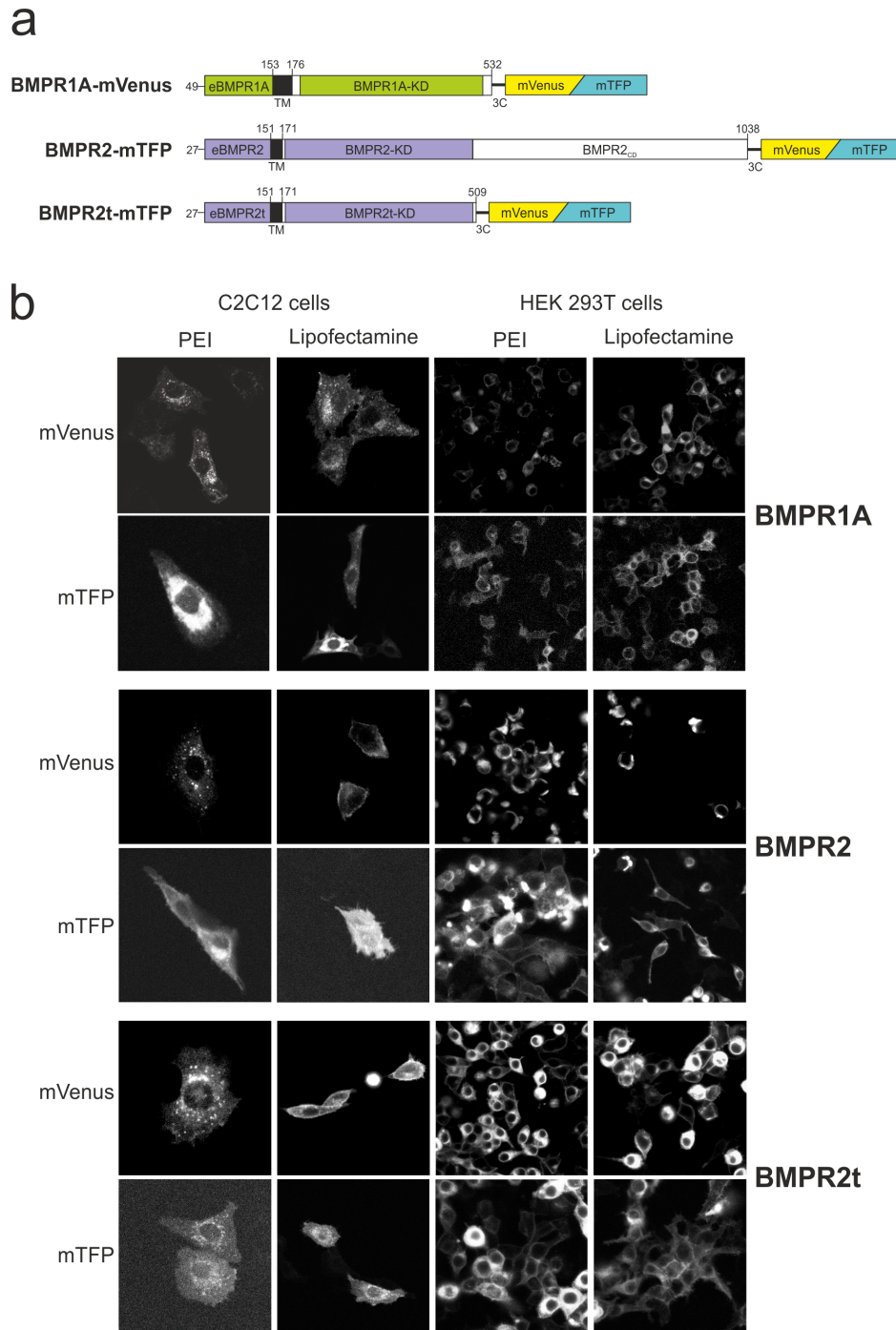


Figure 4.20: Optimisation of fluorescently tagged BMP-receptor transfection. (a) Schematics of the fluorescently tagged BMP receptor constructs BMPR1A, BMPR2t (truncated) and BMPR2 (full-length). BMPR1A is shown in green, the ectodomain and kinase domain (KD) of BMPR2 are shown in purple. The C-terminal of BMPR2 (BMPR2_{CD}) is shown in white. TM = transmembrane helix; 3C = 3C protease linker. The fluorophores mTFP (cyan) and mVenus (yellow) are attached via a linker to the receptors C-termini. (b) Expression tests of BMPR1A-mVenus/mTFP, BMPR2-mVenus/mTFP and BMPR2t-mVenus/mTFP in C2C12 and HEK293T cells using either PEI or Lipofectamine[®] 2000 transfection reagents. Representative fluorescence microscopy images are shown.

In an initial FRET-FLIM experiment, COS7 cells were transfected with BMPR2-mTFP, BMPR2t-mTFP, BMPR2-mTFP/BMPR1A-mVenus or BMPR2t-mTFP/BMPR1A-mVenus and the average donor (mTFP) fluorescence lifetime was measured before and after addition of BMP2 (Fig. 4.21). A decrease in donor lifetime correlates with FRET between the mTFP and mVenus fluorophores, which indicates complex formation between the two receptors. For BMPR2-mTFP ($n = 15$) and BMPR2t-mTFP ($n = 15$) alone, the average lifetimes were similar, 2.50 ns and 2.54 ns respectively. The average lifetime for cells transfected with BMPR2-mTFP and BMPR1A-mVenus ($n = 9$) decreased slightly to 2.44 ns ($P = 0.5828$), but upon addition of BMP2 the average lifetime showed a significant decrease ($n = 15$, $P < 0.0001$) to 2.16 ns, demonstrating that the ligand induced the formation of receptor-receptor complexes (Fig. 4.21). For cells transfected with both BMPR2t-mTFP and BMPR1A-mVenus ($n = 7$) the average lifetime was 2.30 ns, which was a significant reduction from that for BMPR2t-mTFP alone ($P < 0.0001$). Following the addition of BMP2 ($n = 8$) there was then only a small further reduction in lifetime to 2.20 ns, which was not significant ($P = 0.4944$) (Fig. 4.21). From this analysis BMPR2-mTFP was selected for subsequent FRET-FLIM experiments. The decrease in fluorescence donor lifetime in BMPR2t-mTFP/BMPR1A-mVenus co-transfected cells was much larger than that for BMPR2-mTFP/BMPR1A-mVenus and this prevented the observation of a significant decrease upon BMP2 addition. This is probably due to the longer cytoplasmic domain of the full-length receptor increasing the distance between the mVenus and mTFP fluorophores such that only when the receptors are tightly associated in ligand-induced dimers are the two fluorophores close enough for FRET to occur. Since the location of BMP-induced receptor complexes was the focus of this study BMPR2t-mTFP was not useful in further experiments. Interestingly, this analysis confirms the existence of pre-formed BMP receptor complexes that are thought to be important in regulating BMP signalling[186, 187].

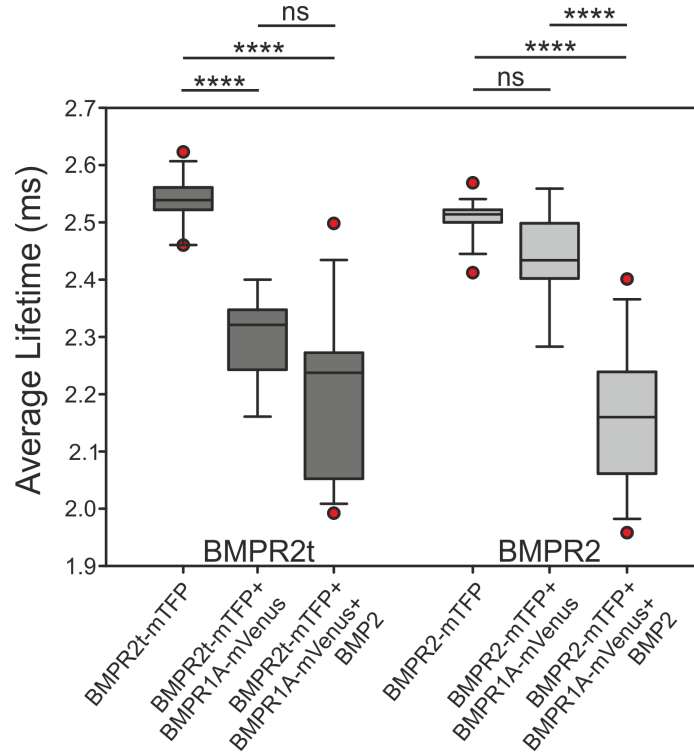


Figure 4.21: BMPR2 is most suitable for FRET FLIM experiments.

Box plot showing average donor lifetimes (ns) recorded in FRET FLIM experiments of cells transfected with the indicated fluorescently-tagged BMP receptor constructs and with or without addition of BMP2. The boxes indicate the 25th to the 75th percentiles; the median is indicated by a line within the box; the whiskers mark the 10th and 90th percentiles; points outside of these limits are shown as red dots. Data are from one independent experiment (n = 15, 5, 8, 15, 9, 15 left to right); **** $P < 0.0001$, ns = 0.4944 (BMPR2t), 0.5828 (BMPR2) by one-way ANOVA with Tukey multiple comparison test.

To determine the subcellular localisation of BMP2-induced BMP receptor complexes, COS7 cells were transfected with either BMPR2-mTFP alone or both BMPR1A-mVenus and BMPR2-mTFP. Subsequently, live-cell quantitative FRET analysis using fast acquisition time domain FLIM[277] was performed before and after addition of BMP2 (Fig. 4.22). This technique requires shorter acquisition times and therefore yields a high spatial resolution of FRET events within living cells. Using this technique, pixel-by-pixel measurements of the donor (mTFP) lifetime and the minimal fraction of donor molecules undergoing FRET (f(D)) were obtained (Fig. 4.22a). The addition of BMP2 did not affect the average donor lifetime in cells transfected with BMPR2-mTFP alone (Fig. 4.22a,b). In

some cases clustering of BMPR2-mTFP within the cell was observed in the intensity images after the addition of BMP2, which could potentially be due to internalisation of the receptor in the presence of the ligand.

Cells transfected with both BMPR1A-mVenus and BMPR2-mTFP showed a slight decrease in the average lifetime (Fig. 4.22b). The pixel-by-pixel analysis showed that this interaction seemed to occur in a diffuse pattern across the cell membrane (Fig. 4.22a). In contrast, after addition of BMP2, a marked decrease in the average donor lifetime was observed (Fig. 4.22b). The pixel-by-pixel analysis showed an increased clustering of BMPR2-mTFP in a punctate pattern within the cell, and these puncta correlated with a decreased donor lifetime and higher $f(D)$ values (Fig. 4.22a). It is tempting to speculate that this shows that upon addition of ligand there is an increase in receptor-receptor FRET specifically within the endosomes. However, due to the lack of an endosomal marker protein it is not certain that these puncta represent endosomes. Furthermore, these data proved difficult to replicate and it was not possible to observe enough cells to generate robust statistics for analysis of the average fluorescence lifetimes. The difficulties ranged from finding BMP-responsive cells (i.e. cells in which we observed a decrease in lifetime upon addition of ligand), to laser damage that affected cell morphology, to difficulties optimising the ratio between BMPR1A-mVenus and BMPR2-mTFP.

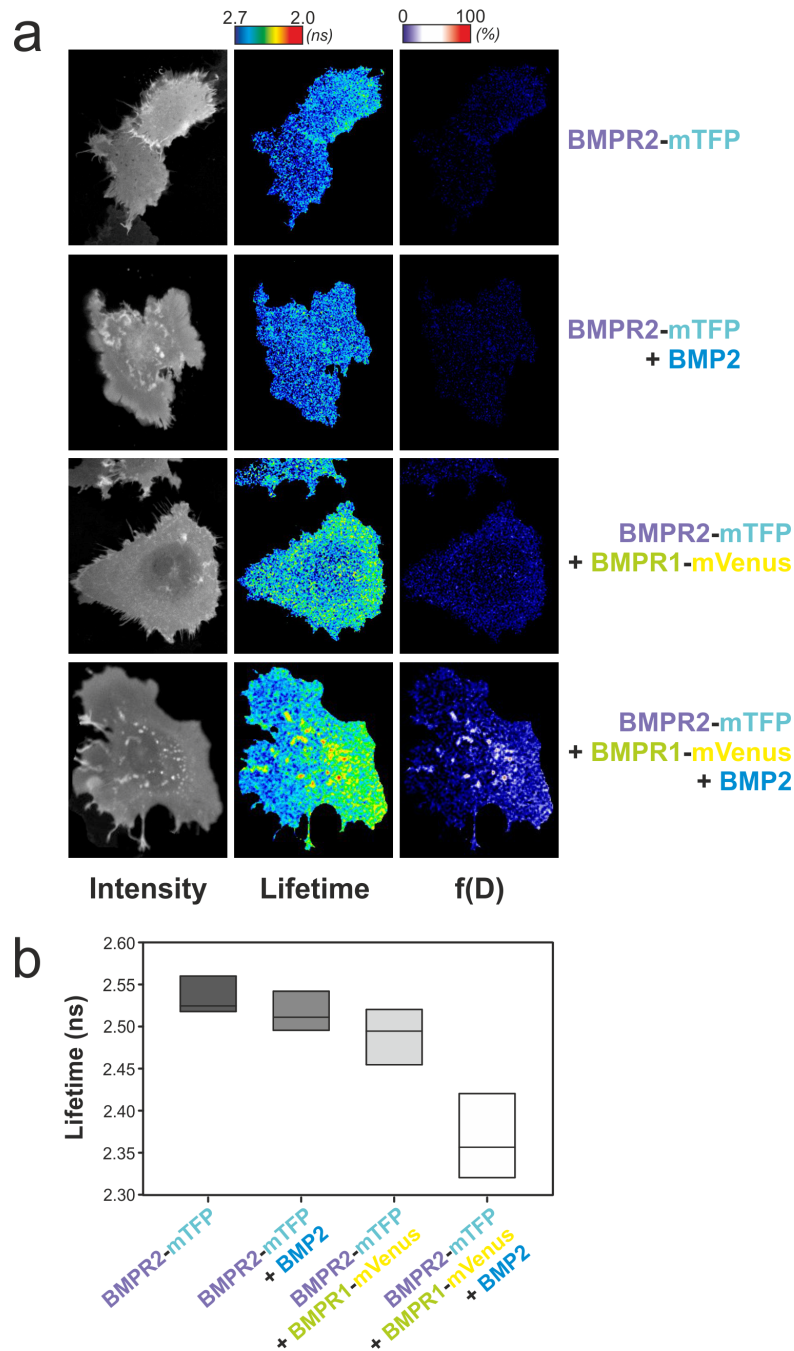


Figure 4.22: Spatial analysis of BMP receptor FRET-FLIM data. (a) Images of representative cells transfected with the indicated fluorescently-labelled BMP receptors +/- BMP2. BMPR2-mTFP intensity images (left column), pixel-by-pixel analysis of donor lifetime (from 2.7 ns, blue to 2.0 ns, red) (middle column) and the fraction of interacting donor (fD) (from 0 %, blue to 100 %, red) (right column) are shown (Symphotime, ImageJ)^[277]. (b) Box plot of mean cell donor lifetime (ns) for each condition. Boxes indicate the 25th to the 75th percentiles and the central line indicates the median. Data are from one independent experiment (n = 5, 5, 6, 5 left to right), full statistics could not be calculated for the data as not enough replicates were performed.

4.3.11 Time-resolved analysis of BMP receptor FRET

Following on from the spatial analysis of BMP receptor FRET, in order to examine the effects of RGM on the localisation and rate of formation of active BMP-induced receptor complexes, further FRET-FLIM experiments were performed. COS7 cells were co-transfected with BMPR1A-mVenus/BMPR2-mTFP or BMPR1A-mVenus/BMPR2-mTFP/RGMB (GPI-anchored). Several FRET-FLIM acquisitions were performed over a time-course spanning the addition of BMP2, and the pixel-by-pixel lifetime of BMPR2-mTFP was recorded (Fig. 4.23a). For each cell within the field of view, histograms from each time point were assembled reporting the average donor lifetime for each pixel within the cell (Fig. 4.23b). A graph of the mean donor lifetime for each cell at each time point was plotted, which showed changes in the donor lifetime in response to the addition of BMP2, in the presence and absence of RGMB (Fig. 4.23c). The mean lifetime was recorded several times prior to the addition of BMP2 to demonstrate that the donor lifetime was stable and not affected by other variables such as laser damage. A marked decrease in the average donor lifetime was observed upon addition of BMP2, but this did not seem to be significantly different in cells also expressing RGMB (Fig. 4.23c). However, during this analysis it was found that not all cells were BMP-responsive (i.e. did not show a decrease in average donor lifetime in response to BMP2), and those that were showed great disparity in the donor lifetimes recorded (Fig. 4.23a,b). For example, representative histograms show that the initial donor lifetime before addition of ligand varied markedly between cells, and both the magnitude and rate of the reduction in lifetime following addition of BMP2 was also diverse (Fig. 4.23b). The result of this natural variation in the BMP-responsiveness of cells resulted in a large standard deviation of the lifetime values at each time point. Consequently, a significant difference between RGMB-transfected and control-transfected cells could not be recorded. Unfortunately, it was therefore not possible to collect any

statistically robust microscopy data in support of the endocytosis-linked model, however, this work is in its early stages and with further optimisation these experimental setups have the potential to yield insights into the mechanism of RGM-activated BMP signalling.

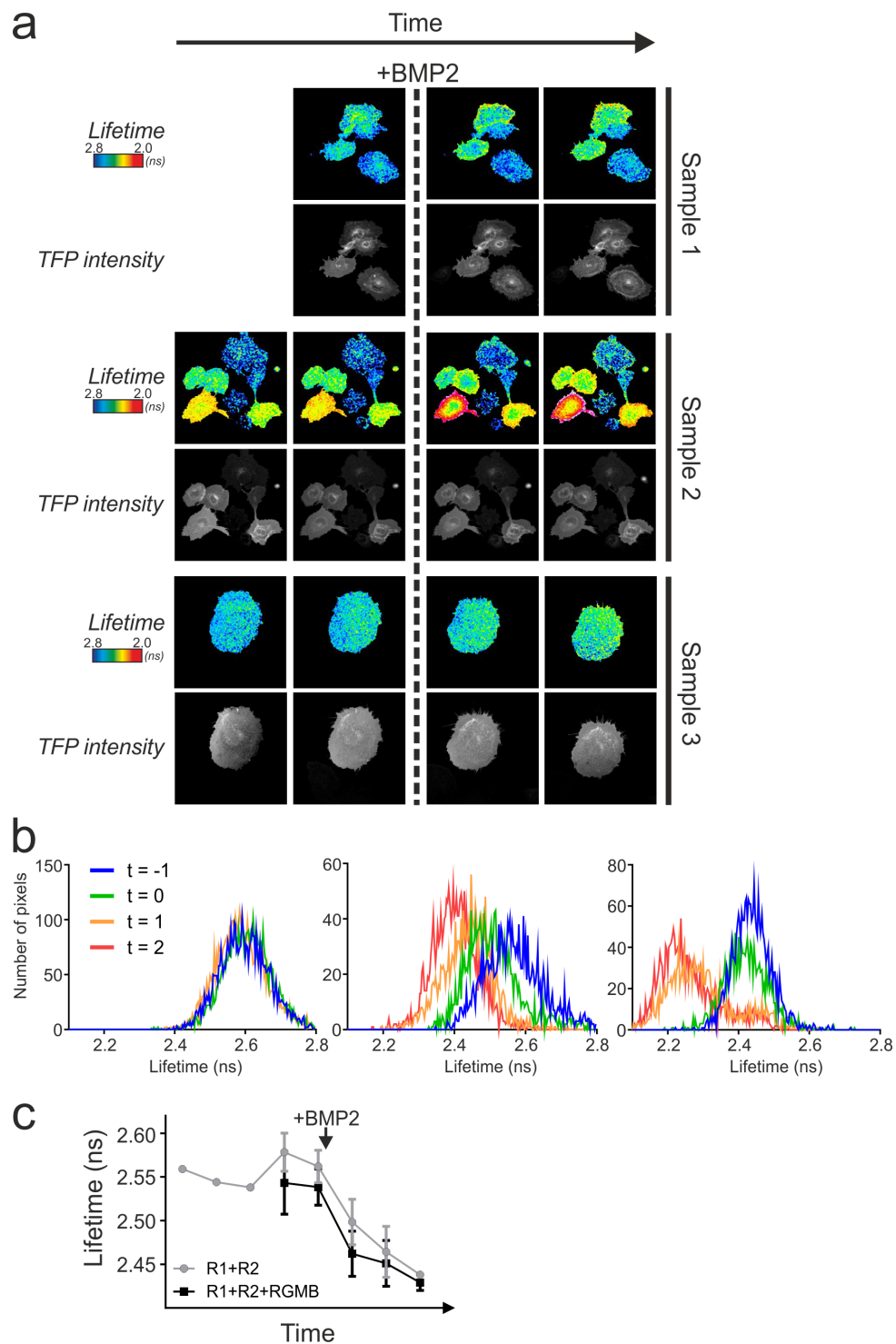


Figure 4.23: An illustration of the difficulties associated with the FRET-FLIM experiments. (a) Images of COS7 cells transfected with BMPR2-mTFP and BMPR1A-mVenus taken from three independent time courses during which BMP2 was added (indicated by a dotted line). FRET-FLIM data were acquired at each time point. BMPR2-mTFP intensity images (second row) and the corresponding fluorescence donor lifetime pixel-by-pixel analysis, artificially coloured (from 2.8 ns, blue to 2.0 ns, red) heat-map images are shown (first row). (b) Histograms of pixel-by-pixel lifetimes from three different cells over a time course, $t = -1$ (blue curves), $t = 0$ (green curves), $t = 1$ (orange curves) and $t = 2$ (red curves); $t = 0$ corresponds to the FLIM acquisition immediately after BMP2 addition. (c) A plot of average fluorescence donor lifetime over time for cells transfected with either BMPR2-mTFP (R2) + BMPR1A-mVenus (R1) (data shown in grey; $n = 1, 1, 10, 13, 13, 13, 1$; where n is number of cells analysed at each time point) or R2 + R1 + RGMB (data shown in black; $n = 0, 0, 0, 9, 16, 16, 16, 2$). The time point at which BMP2 was added is indicated. Error bars are s.e.m. and data are from three independent experiments.

4.4 Discussion

4.4.1 An analysis of RGM interactions

A thorough analysis of the interactions between different constructs of all three RGM family members with both BMP2 and the BMP receptors was carried out (Fig. 4.13 and 4.15). Interestingly the fit of the equilibrium binding data in all of the SPR analysis involving the full-length RGM proteins showed a biphasic curve progression, indicating underlying additional non-specific binding. This led to the use of a “total binding model” for data fitting, which accounts for a 1:1 binding taking into account additional non-specific interactions. A comparison of the fits for the total and specific binding models to the eRGMC-BMP2 SPR equilibrium binding data justified the use of a model that incorporated non-specific interactions (Fig. 4.12a). The use of this model was further validated by kinetic analysis of SPR binding data for the interactions of RGMB_{ND} and eRGMB with BMP2 (Fig. 4.12b,c). Whilst a good fit was obtained for the RGMB_{ND} data (Res. sd = 2.456), the 1:1 Langmuir binding model did not fit the

eRGMB data (Res. $sd = 21.321$). This highlights the extent of the non-specific binding between eRGMB and BMP2, which could potentially be attributed to the middle domain of RGM where, opposite the NEO1-binding face (Fig. 3.2), there is a large cluster of hydrophobic residues that might be responsible for the sticky behaviour of the protein. One can speculate that *in vivo* this portion of the protein might be blocked, perhaps by another binding partner.

The binding affinities calculated from the SPR analyses were between 50 to 100-fold lower than some published studies of RGM interactions with BMP ligands[302]. For example for eRGMB binding to BMP2 our SPR analysis yielded a K_d of 400 nM whereas some published values are around 5-10 nM. The reason why the binding affinities from this work are so different to the published values is not clear, however, all RGM proteins used in this work were produced in a well-established mammalian expression system[200] resulting in high-quality monodisperse proteins with a purity of >95 % (Fig. 4.3c). Furthermore, the BMP2 was produced using an established refolding protocol[206] and purified to >95 % dimeric species, and was shown to be highly active in luciferase and SMAD-phosphorylation assays (Fig. 4.5). Moreover, it is worth noting that the affinities measured for the eRGMC-BMP2 interaction are similar to those reported in another study from the Bjorkman laboratory[280]. They report an affinity of $K_d = 140$ nM for this interaction, which is similar to the value obtained in this work for the eRGMC-BMP2 binding analysis using a 1:1 Langmuir binding model (Fig. 4.12a). Interestingly the SPR binding data for the eRGMC-BMP2 interaction from the Bjorkman laboratory shows a similar biphasic curve progression to that shown here.

An analysis of the RGMB_{ND}-BMP2 interaction was also carried out using MALS, and this showed that the formation of the complex is dependent on pH (Fig. 4.17a). The motivation for this experiment was a result of identifying H106^{RGMB} as a key RGMB-BMP2 interface residue and hypothesising that protonation of

this residue might disrupt complex formation. However, a recent paper studying the interaction energies of different forms of histidine using quantitative quantum chemical calculations showed that the π - π stacking interaction of neutral histidine with aromatic amino acids (Phe, Tyr, Trp) has an energy in the range of -3.0 to -4.0 kcal/mol[309]. This interaction is much stronger than Van der Waals forces, explaining why this π -stacking network was found to be important for efficient RGM-BMP2 complex formation (Fig. 4.10c and 4.13h,i). This paper also showed that the π - π stacking interaction energies between protonated histidine and other aromatic amino acids are in the range of -3.6 to -8.3 kcal/mol; which is much larger than that for neutral histidine[309]. The implication of this analysis is that protonation of H106^{RGMB} may actually result in tighter binding, although the specific local environment of His106^{RGMB} within the complex might alter this conclusion. It is possible that other changes at the RGMB-BMP2 interface or in the overall structure of the RGMB protein cause the pH sensitivity of the complex. It would be interesting to investigate the precise basis for the pH dependence of the interaction, perhaps using Nuclear Magnetic Resonance (NMR) studies to analyse the protonation of residues within RGMB_{ND} across a pH titration. This experiment would be feasible given the small size of the N-terminal domain of RGMs (approximately 10.5 kDa).

4.4.2 A model for RGM-activated BMP signalling

The RGM_{ND}-BMP2 structures presented here have revealed for the first time the fold of the RGM N-terminal domain and the mode of RGM interaction with the BMP ligand. These structural data combined with subsequent biophysical and cellular experiments suggest a potential model for RGM-activated BMP signalling that is linked to pH-dependence of the RGMB-BMP2 interaction and the subcellular localisation of the BMP signalling complexes (Fig. 4.24). The

hypothesis is that in the absence of RGMs BMP signalling will take place primarily from the plasma membrane as previously described. However, when RGM is present it will compete with endogenous type I receptors on the cell membrane for binding to the BMP ligand. The RGM-BMP complex may recruit the type II receptor, which, can be accommodated in the architecture of the crystal structure presented here. Following clathrin-mediated endocytosis of the RGM-BMP complex and acidification of the endosomal compartment, RGM will dissociate from BMP, allowing binding of the ligand to an internalised pool of BMP receptors. From this endosomal platform a higher level of BMP signalling may be triggered; both through increased proximity to the nucleus, and through the enhanced availability of downstream effector targets such as SMADs, which may be anchored to the endosome[184]. This model is largely based on both the pH-dependence of the RGM-BMP2 interaction and the contrasting behaviours of eBMPR1A and eRGM in BMP luciferase reporter assays. Several experiments designed to test this model in cells using fluorescence microscopy were performed, including tracking fluorescently labelled BMP ligand and analysing the location of interaction of the BMP receptors, however due to experimental difficulties these did not yield conclusive evidence in support of this model.

The experimental data presented in this chapter have shed light on the mode of action of RGMs in BMP signalling but many interesting questions remain. For example, can BMP type II receptors bind to the RGM-BMP complex? This is currently unclear, although ACVR2A can certainly be accommodated in the architecture of the complex as shown above (Fig. 4.16a). However, whether the complex RGM-BMP2-ACVR2A can form in solution has not yet been tested. This experiment is not trivial because of the low affinity between BMP2 and ACVR2A ($35 \mu\text{M}$, Fig. 4.15b). If the ACVR2-BMP2-RGM complex was able to form, then according to the endocytosis-linked hypothesis, the type II receptor could be internalised indirectly through its interaction with the BMP

dimer. It is possible to create variants of BMP2, such as the double mutant L100K/N102D^{BMP2}, with a high affinity for the Activin type II B receptor (ACVR2B) (K_d 140 nM)[310] and it may be possible in the future to use these to test for the formation of an ACVR2B-BMP2-RGMB complex using MALS.

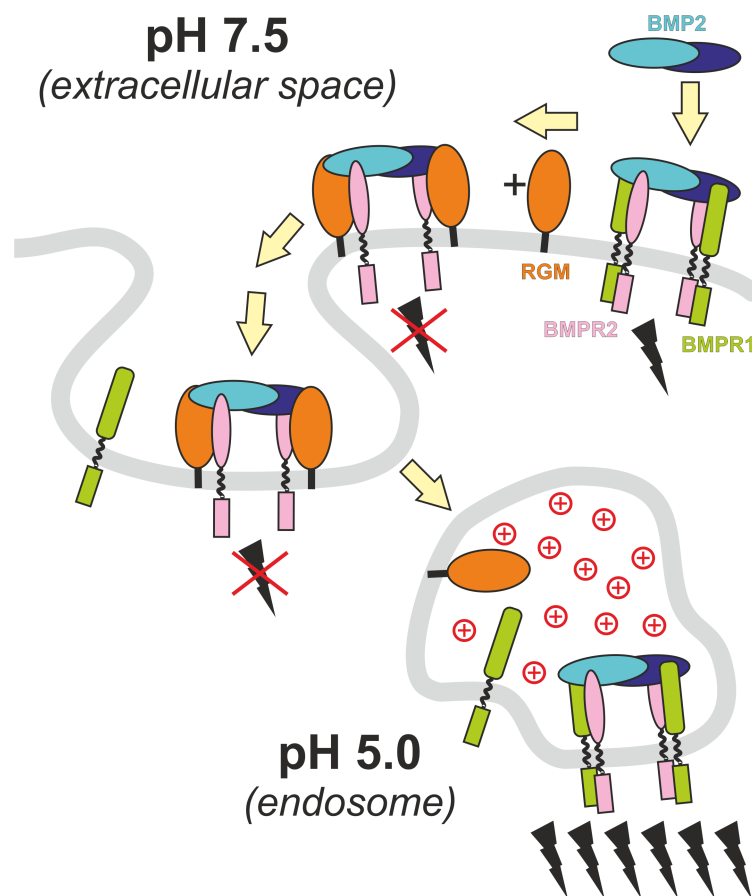


Figure 4.24: Model for RGM-activated BMP signalling. In the absence of RGMs BMP2 binds to the BMP receptors triggering signalling from the plasma membrane. When RGM is present it binds to BMP2 in the neutral pH extracellular environment, potentially also recruiting the type II receptor. This complex is targeted for internalisation and following acidification of the endosomal compartments RGM dissociates from BMP2 allowing the ligand to bind to the internalised pool of BMP receptors. BMP signalling from the endosomal platform is enhanced compared to that triggered from the membrane.

Further to this, it is also not clear what the true arrangement of the RGM-BMP signalling complex would be at the cell membrane in the presence of the BMP receptors. The structure of the BMP ligand bound to two copies of the type I receptor and two copies of the type II receptor has been widely regarded to represent the active signalling complex[298]. However, this has not been investigated *in vivo* and it is possible that in the presence of RGM a heterogenous asymmetric complex might be formed containing a BMP ligand bound to one type I receptor, one type II receptor and one RGM. However, as RGMs do not appear to bind directly to the BMP receptors, and in an SPR competition experiment eRGMB did not enhance but actually competed with binding of the BMPR1A ectodomain to the BMP ligand, formation of this complex seems unlikely.

The proposed model for RGM-activated BMP signalling involves both endocytosis of the RGM-BMP complex and also endocytosis of the BMP type I and type II receptors (Fig. 4.24). The cellular localisation and the site of action of BMP receptors are still not clear. It has been shown that the BMP receptors (type I and II) undergo constitutive clathrin-mediated endocytosis even in the absence of the BMP ligand, resulting in a potentiation of SMAD-dependent BMP signalling upon BMP ligand exposure[180]. Additionally, BMPR2 is also internalised through caveolae and the balance between caveolae- and clathrin-mediated endocytosis was suggested to modulate the patterns of gene transcription initiated by BMP signalling[180, 181]. Therefore, although the BMP receptors are termed “cell-surface” receptors, the actual proportion of receptor on the cell surface at any one time may actually be quite low. Fluorescently tagged forms of the BMP receptors could be used to investigate their subcellular localisation, however, these would best be expressed under the control of the endogenous promoter to avoid overexpression artefacts.

An SPR competition experiment demonstrated that RGMB can effectively compete with BMPR1A binding to BMP2 at neutral pH, a $1.7\times$ molar excess of

BMPRI1A was required to compete with RGMB (Fig. 4.16e). It is not known what the relative expression levels of these two proteins are at the cell surface, and this will certainly vary depending on the cell-type. Understanding this is key to establishing whether this competition is physiologically relevant and may explain why the effects of RGMs often appear to be cell-type specific. It is not necessarily the case that RGM will sequester all of the BMP2 pool at the cell-surface but rather that it may divert a significant proportion of the BMP ligand toward endocytic pathways, resulting in activation of the BMP pathway, although this remains to be demonstrated.

4.4.3 A cellular approach to analysing BMP signalling

For functional analysis of the RGM-BMP interaction a cell-based assay was used to measure BMP signalling via a BRE-luciferase reporter[267]. In the presence of membrane-anchored RGMB, the level of BMP signalling in LLC-PK1 cells stimulated with 6 nM BMP2, was enhanced by approximately 3-fold (Fig. 4.17c), which is in line with data reported in the literature[83, 89]. In similar assays, soluble eRGMB had neither an activating nor an inhibitory effect on BMP signalling, unlike eBMPRI1A, which acted as a strong inhibitor (Fig. 4.17d and 4.18b-d). This result is in contrast to data reported in the literature[311, 89], where soluble RGMB has been shown to attenuate BMP signalling. However, the luciferase assay analysis of the soluble proteins reported in this thesis was performed by transfecting plasmids expressing soluble eRGMB and eBMPRI1A as well as also by adding the proteins directly to the cells; in this variant of the experiment two different cell types, three different concentrations of BMP2 and a wide range of eRGMB and eBMPRI1A concentrations were tested, all yielding similar results (Fig. 4.18b-d). One potential explanation for the differences observed for the behaviour of soluble RGMB could be the use of Fc-RGMB fusion proteins[89].

Not only are Fc fusion proteins dimeric, which will give Fc-RGMB proteins an artificially high affinity for BMP ligands potentially affecting the results of the assay, but furthermore, the dimerisation of RGMB may also block the binding of required protein partners, including, for example, NEO1. It is also worth noting that the cell lines commonly used in these luciferase assays (including HepG2 and LLC-PK1 cells) express Fc-receptors on their surfaces[312, 313]. These may well interact with Fc-RGMB fusion proteins and again affect the results of the luciferase assay. For the assay conducted here, monomeric human RGMB produced from a mammalian cell expression system with a minimal His₆ purification tag was used (Fig. 4.3c), which may account for why soluble eRGMB did not appear to inhibit BMP signalling. In the other study that observed inhibition by soluble RGMB[311], Fc-fusion proteins were not used. However, this study also found full-length membrane-bound RGMB to be an inhibitor of BMP signalling, which suggests that both results may have been due to general differences in the protocol used.

Although the data supporting soluble RGMB acting as an inhibitor of BMP signalling is not conclusive, there is clear evidence for inhibition by soluble RGM[107, 103, 110, 111, 314]. It has been proposed that this might be another level of blood iron level regulation, particularly as soluble RGM has been found in human sera at levels competent to suppress BMP signalling[103, 110, 111]. However, there are some reports contradicting this role for soluble RGM, for example injection of anti-RGM antibodies into rats and monkeys results in a downregulation of hepcidin expression and an increase in serum iron levels[114]; presumably these antibodies bind to circulating soluble RGM and these downstream effects would not be expected if soluble RGM was an inhibitor of BMP signalling. Furthermore, patients with iron-deficiency anaemia have lower circulating levels of soluble RGM than healthy control patients[112], suggesting that increasing levels of soluble RGM is not a physiological response to con-

trol serum iron levels. It therefore seems that the behaviour of soluble RGM proteins is possibly cell-type, RGM family member and context specific. The model for RGM-activated signalling proposed here is therefore probably vastly oversimplified.

A series of fluorescence microscopy experiments were used to test this proposed model for RGM-activated BMP signalling, including tracking the internalisation of the BMP2 ligand and analysing the subcellular localisation of BMP receptor FRET. However, although the data obtained for the FRET experiments were promising it proved difficult to reproduce. This was mostly due to vast differences in the level of BMP responsiveness between the cells analysed (Fig. 4.23). It would be interesting to combine the FRET measurements with the use of a cell line containing a stably transfected BMP pathway GFP reporter[315, 316], which would allow the quantification of the level of activation of the BMP pathway. Non- or hyper-responsive cells could then be excluded from our analysis and the correlation between fluorescence lifetime values and BMP pathway activation could be confirmed. Including a reporter for non-SMAD signalling, activating expression of a different fluorescent protein, would also allow the ratio of SMAD versus non-SMAD signalling initiated by BMP2 to be analysed in the presence and absence of RGMs.

4.4.4 Implications of the RGM-BMP2 complex structure for JHH

Importantly, the RGM_{ND}-BMP2 structures provide a rationale for the molecular mechanism of JHH-linked mutations in RGMC. JHH is an autosomal-recessive iron overload disorder that results in cardiomyopathy and diabetes. JHH is caused by a deficiency in the levels of hepcidin, whereas an excess of hepcidin is linked to anaemia of inflammation[99, 101]. RGMC-activated BMP signalling is

crucial for up-regulation of hepcidin expression and control of serum iron level, and mutations in RGMC are the major cause of JHH[102, 54]. Most of these mutations are located in the C-terminal domain of RGMC, the region responsible for NEO1 binding (see chapter 3)[240]. These mutations impair protein secretion in mammalian cells, whereas mutations located in the RGMC N-terminal domain do not affect secretion[240, 317]. The RGMC_{ND}-BMP2 crystal structure revealed that two of these mutations, G99R^{RGMC} and L101P^{RGMC} are located directly at the interface with BMP2 and reduce the affinity of the RGMC-BMP2 interaction (K_d 910 nM, 1.5 μ M respectively, from 124 nM) (Fig. 4.13j,k). This not only validates the interface observed in the RGM_{ND}-BMP2 structures, but also suggests that disruption of the BMP6-RGMC interaction is the molecular mechanism for JHH disease pathology. Taken together, our analysis may provide a basis for the structure-guided design of novel therapeutics for the treatment of iron related disorders such as hemochromatosis and anaemia of inflammation.

4.5 Summary

In conclusion, this chapter details the solution of the first structure of the RGM N-terminal domain in complex with the BMP2 ligand, which has informed a molecular mechanism of JHH-linked mutations in RGMC. This structural data, together with biophysical and cellular analyses have allowed a potential mechanism for RGM-activated BMP signalling to be proposed. Further work will be required to test this model, for example demonstrating the involvement of RGMs in BMP ligand endocytosis and showing that the endosome functions as a platform for BMP signalling. It will be interesting to see whether all RGM family members act via the same mechanism or whether the biological context, including the available protein binding partners, play a large role in directing how RGMs affect the BMP pathway. However, translocation of the signalling

machinery through established pathways to place it in close proximity to the nucleus, and thus downstream effector targets, is a very efficient way of effecting changes in gene expression. This mechanism has been suggested for other signalling pathways including the closely related $\text{TGF}\beta$ [179], epidermal growth factor (EGF)[318], and glial-cell derived neurotrophic factor (GDNF) signalling pathways[319]. The molecular mechanism of RGM-mediated BMP activation, based on endocytosis of the entire signalling complex may provide a paradigm for many “cell-surface” signalling events.

Chapter 5

RGM is a structural bridge
between BMP and NEO1
signalling.

5.1 Abstract

There are many lines of evidence suggesting that the NEO1 and BMP signalling pathways are functionally linked, most likely through the actions of RGMs[199, 190, 191, 197], the most compelling being the iron overload observed in the livers of NEO1^{-/-} knockout mice[160]. However, a physical demonstration of RGM acting as a link between these two signalling pathways has not yet been reported. In this chapter I present a high-resolution crystal structure of the ternary BMP2-eRGMB-NEO1_{FN56} complex, together with a thorough interface analysis and SAXS experiments demonstrating the nature of this complex in solution. In addition, superposition of the ternary complex with the active signalling conformation of the eRGMB-NEO1_{FN56} suggests a model by which RGMs mediate clustering of BMP and NEO1 dimers on the plasma membrane. This model is further supported by TIRF-dSTORM microscopy experiments, which demonstrate a time-dependent clustering of fluorescently labelled NEO1 receptors upon addition of BMP2 in the presence of RGMB. Together these data show that RGM is a structural bridge between NEO1 and BMP signalling, and inform a clustering mechanism that may be important in the activation of these signalling pathways.

5.2 Introduction

The best-characterised function of NEO1 is as a receptor for RGMs, mediating axon guidance functions[141]. RGM binding to the FN5 and FN6 domains of NEO1 activates downstream signalling through small GTPases of the Rho family, which triggers actin cytoskeleton rearrangements and subsequent axon growth cone collapse. As described in Chapter 3, this effect is mediated by the RGM middle domain, which adopts a tight β -sandwich fold stabilised by an autocat-

alytic cleavage site, and acts as a molecular staple bringing together and fixing the orientation of two NEO1 receptors to activate downstream signalling[240]. RGMs also have another function, as described in Chapter 4, as an important co-receptor for BMP signalling, which is mediated through the binding of the RGM N-terminal domain directly to the BMP ligand[320]. Recent data have probed the relationship between NEO1 and BMP signalling and the potential role of RGM in linking these two pathways.

5.2.1 NEO1 and BMP signalling are physiologically linked

A wealth of recent evidence has shown a physiological link between NEO1 and BMP signalling. In NEO1^{-/-} knockout mice iron overload in the liver is observed[160], a similar phenotype to that observed in RGM^{-/-} and SMAD4^{-/-} conditional knockouts[106, 104]; demonstrating that NEO1 plays a role in the process of iron homeostasis, a function usually attributed to RGM^{-/-}-activated BMP6 signalling[102]. NEO1-deficient mice also show defects in another pathway traditionally thought to be under the control of BMP signalling, chondrogenesis and endochondral ossification; suggesting that NEO1 enhances BMP signalling in chondrocytes[197]. The exact mechanism through which NEO1 enhances BMP signalling is not understood, however, the interaction appears to involve precise spatial regulation of the signalling components into membrane microdomains. NEO1 appears to enhance the association of the BMP receptors with particular microdomains from which BMPs activate downstream SMAD signalling, in a process dependent on RGM[197]. Reciprocally, NEO1 is targeted to lipid rafts in the tips of axon growth cones, a process that is necessary for growth cone collapse and was shown to be dependent on both RGM and BMP signalling[198]. The precise localisation of these signalling molecules appears then to be important for the interdependence of NEO1 and BMP signalling.

The two pathways appear to be linked through the actions of RGMs, although this is still debated by some groups[321, 192]. For example, in *C. elegans*, the up-regulation the Sma/Mab BMP-like pathway requires UNC40, a NEO1 homolog, and DRAG-1, an RGM homolog[199]. Furthermore, in mammalian cells, disruption of the NEO1-RGM interaction using a NEO1_{FN56} fragment abolished NEO1 activation of BMP signalling; similarly, the JHH mutant G99V, which disrupts the RGM-BMP interaction also prevented enhancement of BMP signalling by NEO1[191]. This suggests that the interaction of RGMs with both BMP and NEO1 is required for the crosstalk between the two pathways. In support of this, the most common JHH-linked RGMC mutation G320V, which is the only JHH-linked mutation mapped to the middle domain of RGMC that is secreted, disrupts the RGMC-NEO1 interaction[190]. The corresponding residue in RGMB was shown to be located close to the RGM-NEO1 site-1 interface, with the mutation likely to disrupt the interactions within the binding site[240] (Fig. 5.1). This suggests that uniquely among the JHH-linked RGMC mutations, the disease-causing mechanism of this mutation is disruption of RGMC-NEO1 binding, suggesting that this interaction is important for RGMC-activated BMP signalling.

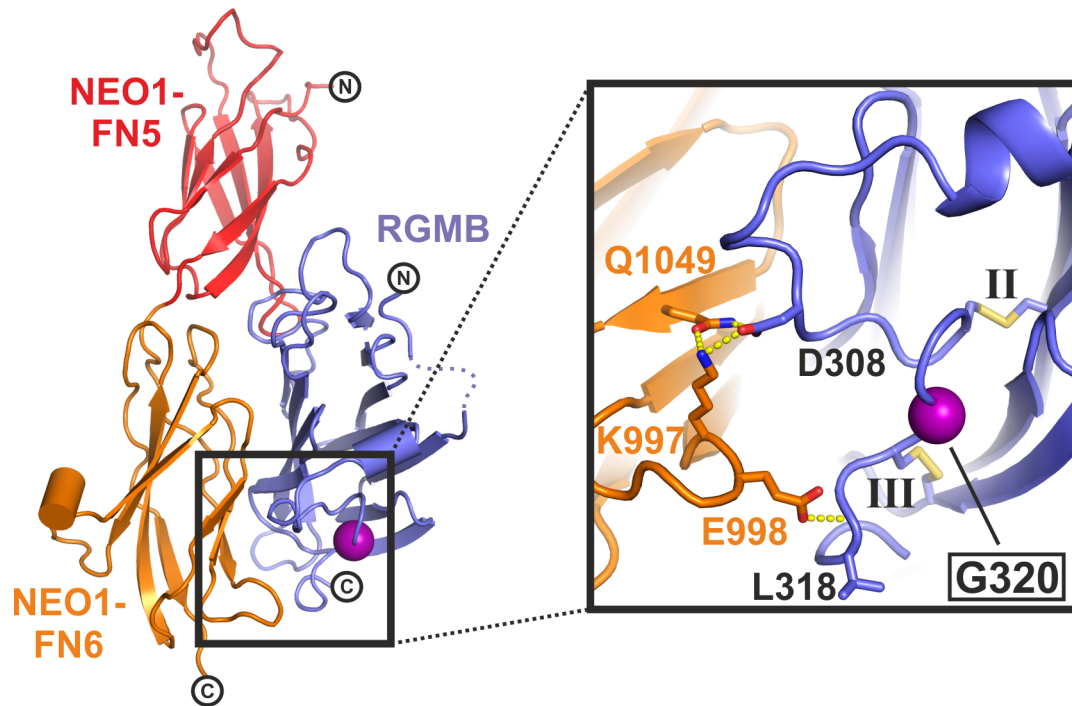


Figure 5.1: The JHH-linked RGM C G320V mutation is located at the NEO1 binding site. From [240], *Science*, reprinted with permission from AAAS. Cartoon representation of the eRGM-NEO1_{FN56} site-1 interface, NEO1_{FN5} is shown in red, NEO1_{FN6} in orange and eRGM in purple. The site of the G320V mutation is indicated using a magenta sphere. The right panel shows a close-up view of the binding interface adjacent to residue G320 (magenta sphere). Disulphide bonds are labelled using Roman numerals, interacting residues are shown as sticks and labelled, and hydrogen bonds are shown as dotted yellow lines. Figure adapted from [240].

5.2.2 RGM plausibly forms a physical link between NEO1 and BMP2

Although many studies have found that RGMs are required for the BMP-NEO1 interaction one study found that NEO1 could directly bind to BMP ligands 2, 4, 6 and 7, with a K_d of 25-50 nM, in co-immunoprecipitation and enzyme-linked immunosorbent (ELISA) assays [321]. Potentially, NEO1 may bind concomitantly to BMP and RGM, thereby enhancing their interaction and thus activating BMP signalling, this remains to be tested. In Chapter 3 and Chapter 4 structural studies of the RGM proteins including the high resolution crystal structures of the

RGMB middle domain in complex with the NEO1 FN5 and FN6 domains, and the N-terminal domains of all RGM family members in complex with BMP2 were presented (Fig. 5.2). These structural data have suggested that the RGM middle and N-terminal domains function independently of each other. In the eRGMB-NEO1_{FN56} complex, despite the eRGMB crystallisation construct comprising the full-length ectodomain of the protein, the N-terminal domain was not resolved in the structure and the N-terminus of the middle domain protruded into a solvent channel in the crystal[240]. Similarly, in the RGM_{C_{ND}}-BMP2 complex the C-terminus projects away from the RGM triple helix bundle, an orientation that is fixed by a disulphide bond[320]. Given this, and the nanomolar binding affinities of both the RGM-NEO1 and RGM-BMP interactions, the two RGM domains might be able to bind to BMP2 and NEO1 simultaneously, thereby bridging the signalling pathways. Whether there is any interaction between BMP and NEO1 in this hypothetical complex remains to be seen. Although RGMs have been implicated in the interplay between NEO1 and BMP signalling, a physical demonstration of the existence of this ternary complex is so far lacking.

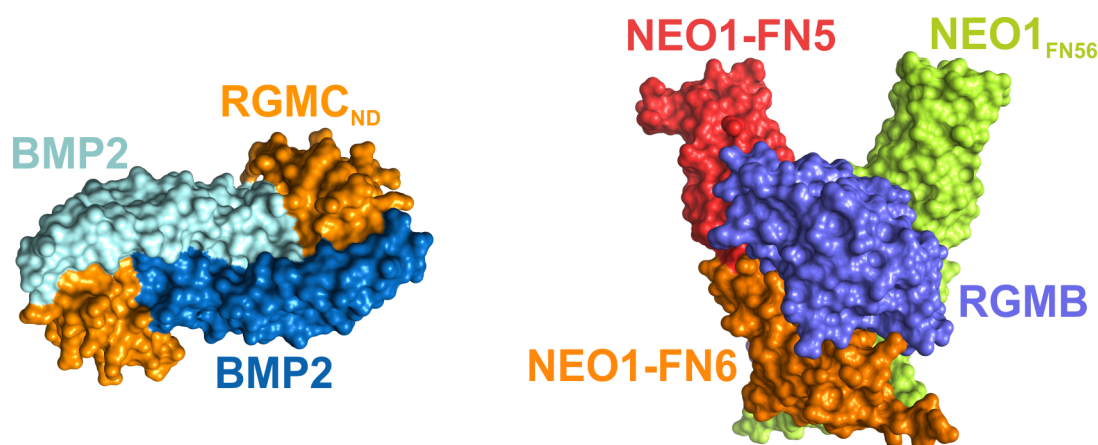


Figure 5.2: The structures of the RGM-NEO1 and RGM-BMP complexes. Surface representations of (left panel) the RGM_{C_{ND}}-BMP2 complex (described in Chapter 4) and (right panel) the active signalling conformation of the eRGMB-NEO1_{FN56} complex[240] (described in Chapter 3). In the left panel RGM_{C_{ND}} is shown in orange and BMP2 is shown in blue/cyan. In the right panel NEO1_{FN5} is shown in red, NEO1_{FN6} in orange, NEO1_{FN56} in green and eRGMB in purple.

5.2.3 Chapter outline

The work presented in this chapter aims to demonstrate that RGM acts as a physical bridge between the NEO1 and BMP signalling pathways. The purification, crystallisation and structure solution of the BMP2-eRGMB-NEO1_{FN56} ternary complex is described. An analysis of interfaces in this complex revealed a small BMP2-NEO1_{FN56} contact and the role of this was analysed and will be discussed. SAXS experiments demonstrated that the ternary complex exists in solution and that the architecture observed in the crystal structure is largely preserved. The crystal structure suggests a model by which the ternary complex can exist as clusters on the cell membrane. This was tested using TIRF-dSTORM imaging in both live and fixed cells using fluorescently tagged full-length NEO1. These experiments showed that BMP2 induces time-dependent clustering of NEO1 receptors on the cell membrane, a process that was dependent on RGMB. This work provides the first demonstration of the existence of a ternary BMP-RGM-NEO1 complex and also informs a potential model by which, through RGM, the interplay between NEO1 and BMP could result in enhanced signalling from both pathways.

5.3 Results

5.3.1 The structure of the BMP2-eRGMB-NEO1_{FN56} ternary complex

NEO1_{FN56} and eRGMB were purified following expression in mammalian cells using Talon-affinity purification and SEC; and BMP2 was refolded from bacterial inclusion bodies, as previously described (Material and Methods section 2.3). The three proteins were then mixed in an estimated 1:1:1 molar ratio and

the complex was concentrated and purified by SEC (Fig. 5.3a). Three peaks were observed on the SEC trace and their identities were confirmed by reducing SDS-PAGE (Fig. 5.3b). Peak III with the lowest molecular mass appeared to consist primarily of excess eRGMB, which ran as two separate bands (~ 30 kDa and ~ 20 kDa) on the reducing gel as expected due to the autocatalytic cleavage site; Peak II appeared to be excess eRGMB-NEO1_{FN56} complex and Peak I was composed of roughly equimolar amounts of NEO1_{FN56}, eRGMB and BMP2 suggesting that these components formed a complex that held together throughout an SEC experiment. The SEC Peak I fractions were pooled for crystallisation trials and crystals grew in one condition at room temperature (Fig. 5.3c). This crystal was tested at DIAMOND, Beamline I03 and diffracted to 3.1 Å, with relatively large unit cell dimensions (Fig. 5.3d, Table 1). The structure of the complex was solved by molecular replacement using the eRGMB-NEO1_{FN56}[240] and RGMB_{ND}-BMP2 (Chapter 4.3.2) structures.

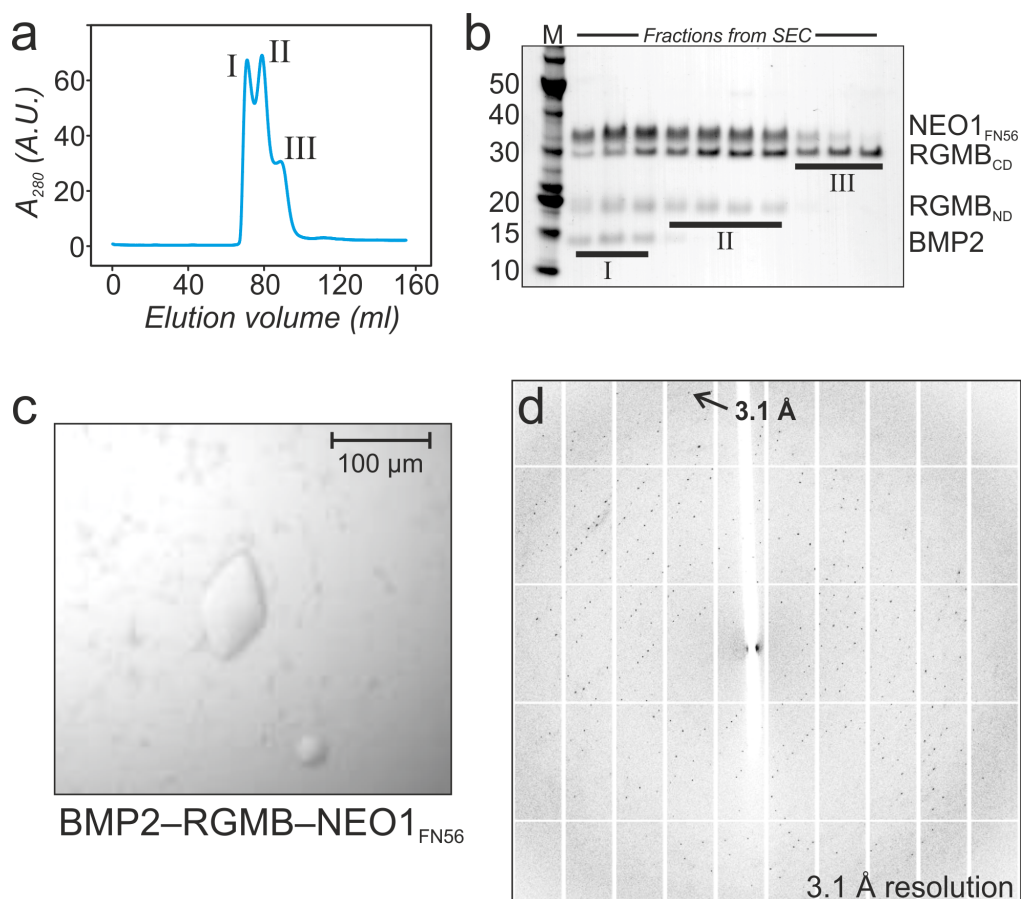


Figure 5.3: Purification and crystallisation of the ternary complex. (a) SEC purification of the ternary BMP2-eRGMB-NEO1_{FN56} complex. The three peaks are labelled I, II and III. (b) SDS-PAGE analysis of the peak fractions from the ternary complex SEC. The fractions corresponding to each peak (I, II and III) are labelled. The bands corresponding to NEO1_{FN56}, RGMC_{CD}, RGMB_{ND} and BMP2 are labelled. M = marker, molecular weights given in kDa. (c) Crystallisation image of the BMP2-eRGMB-NEO1_{FN56} complex from the crystallisation plate, imaged via a Veeco visualisation system[211]. Scale bar indicates 100 μ m. (d) Diffraction image of the BMP2-eRGMB-NEO1_{FN56} complex crystal collected at DIAMOND-I03.

In the complex the disulphide-linked BMP2 dimer binds to two molecules of eRGMB through the RGMB N-terminal domain in a very similar arrangement to that observed in the RGMB_{ND}-BMP2 complex outlined in Chapter 4 (r.m.s.d. of 0.799 Å for 328 equivalent C α positions) (Fig. 5.4). RGMB_{ND} is connected to RGMC_{CD} through a 15-residue linker, which was not resolved in the electron density map and is therefore presumably flexible and disordered. RGMC_{CD} interacts with NEO1_{FN56} through a similar binding mode to that observed for the site-1

interface in the previously determined eRGMB-NEO1_{FN56} complex[240] (r.m.s.d. of 0.511 Å for 368 equivalent C α positions) (Fig. 5.4). The binding modes from the two binary complexes are therefore largely conserved in the ternary complex, however, an additional NEO1_{FN5}-BMP2 interface was also observed (Fig. 5.4 and Fig. 5.5). In the complex the C-termini of NEO1 and RGMB, which, *in vivo*, would be membrane tethered, are oriented in the same direction and the number of amino acids unresolved in the structure provides enough space between the C-terminal domains and the hypothetical membrane attachment site to make the architecture of the complex feasible with respect to the cellular context.

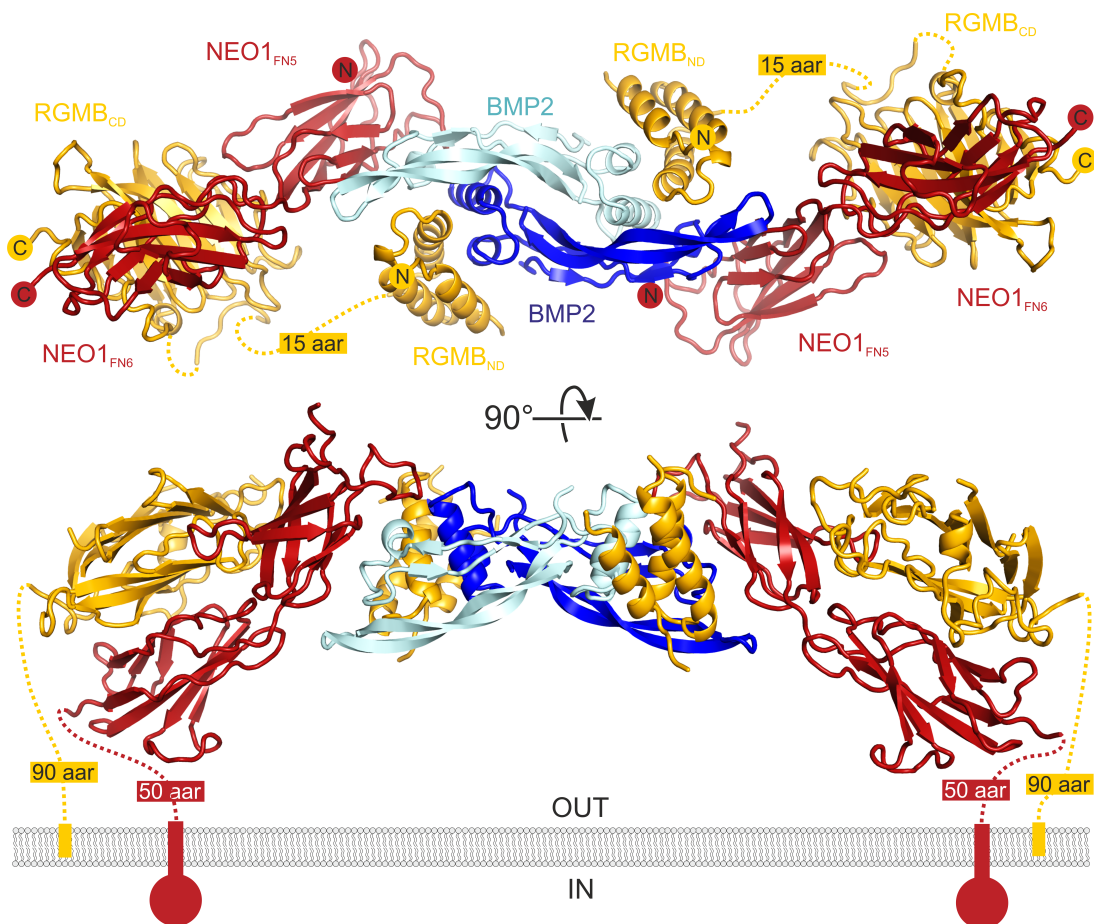


Figure 5.4: The structure of the BMP2-eRGMB-NEO1_{FN56} complex. Reprinted by permission from Macmillan Publishers Ltd: [Nature Structural and Molecular Biology](#), [320] copyright (2015). Cartoon representation of the ternary complex, BMP2 is shown in blue and cyan, eRGMB is shown in gold with RGMBCD and RGMND labelled and NEO1_{FN56} is shown in red. The N- and C-termini of eRGMB and NEO1_{FN56} are indicated. Linkers not visible in the electron density are shown as dotted lines and the number of amino acid residues (aar) are indicated. The orientation of the bottom panel is rotated 90° about the x-axis. A schematic representing the location of the membrane as a phospholipid bilayer (grey) is shown and the predicted protein attachment sites are indicated.

Four interfaces were identified in the ternary complex and to investigate the nature of these, an interface analysis was performed using the programs SC[322] from the CCP4 suite and PISA[254]. The program SC reports the shape correlation statistic, S_c , as a measure of the shape complementarity (SC) of the two surfaces of a protein-protein interface. The software PISA performs a comprehensive analysis of the protein-protein interface, determining parameters such

as the total buried surface area, the solvation free energy gain (ΔG^{int}) and the number of different types of interactions between residues at the interface. All of these parameters are collated and an overall complexation significance score (CSS) is calculated which gives an indication of the importance of the interface for the formation of the complex. These statistics were calculated for each interface in the ternary complex: BMP2-BMP2, RGMB_{ND}-BMP2, RGMCD-NEO1_{FN56} and NEO1_{FN5}-BMP2. The binary BMP2-BMP2, RGMB_{ND}-BMP2 and RGMCD-NEO1_{FN56} interfaces exhibited high SC and CSS scores combined with large buried surface areas, as expected for physiological interactions. Interestingly, the BMP2-BMP2 interface achieved a relatively low SC score, perhaps suggesting a requirement of the disulphide bond for dimer formation. In contrast, the buried surface area of the NEO1_{FN5}-BMP2 interface is at least 3-times smaller than for the other ternary complex interfaces and the SC score for the binding surface was markedly lower; furthermore, the CSS score implied that this interface was not significant for complex formation. This analysis suggests that the NEO1_{FN5}-BMP2 interface might be a crystallisation artefact, implying that this interaction would not be physiologically relevant.

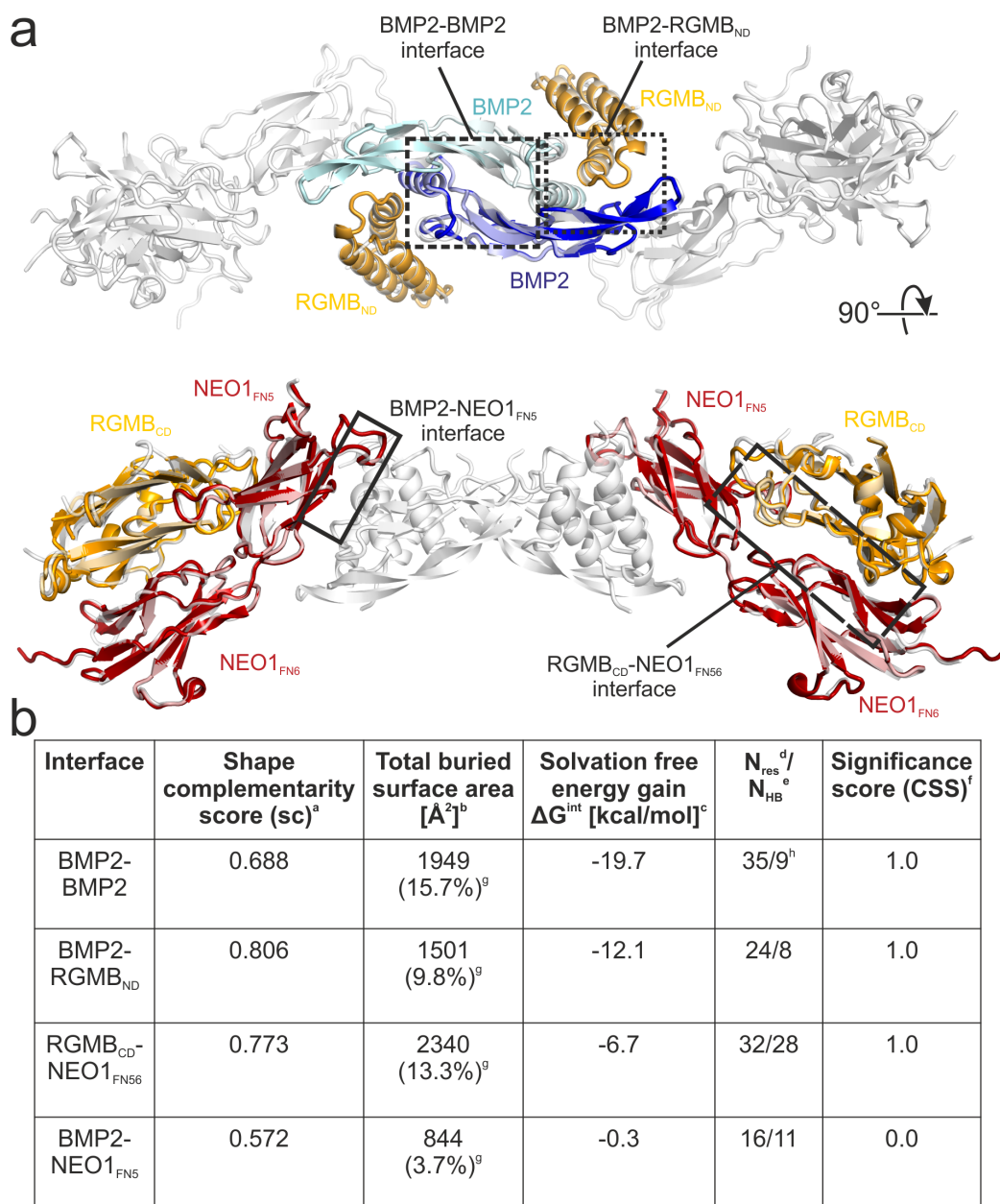


Figure 5.5: Interface analysis and comparison of the ternary BMP2-eRGMB-NEO1_{FN56} complex to the binary eRGMB-NEO1_{FN56} and RGMB_{ND}-BMP2 complexes. Reprinted by permission from Macmillan Publishers Ltd: *Nature Structural and Molecular Biology*, [320] copyright (2015). **(a)** Superposition of the ternary BMP2-eRGMB-NEO1_{FN56} complex (grey) with the binary RGMB_{ND}-BMP2 (gold and blue/cyan) (PDB ID: 4UI0[320]); r.m.s.d. = 0.799 Å for 328 equivalent Cα positions) and eRGMB-NEO1_{FN56} (gold and red) (PDB ID: 4BQ6, site-1 interface[240]; r.m.s.d. = 0.511 Å for 368 equivalent Cα positions) complexes. **(b)** Analysis of the four interfaces highlighted in (a). a = shape complementarity calculated from SC[322]; b = total buried surface area (from PISA[254]); c = solvation free energy gain[254]; d = number of interfacing residues[254]; e = number of potential interface hydrogen bonds [254]; f = Complexation Significance Score[254]; g = percentage overall protein surface buried in the interface; h = excluding the intermolecular disulphide bridge.

To determine whether NEO1_{FN56} binds to BMP2 and whether this interaction enhances RGMB binding, SPR equilibrium binding experiments were performed. No binding of NEO1_{FN56} directly to BMP2 was detected in this experimental setup (Fig. 5.6a). Following SEC purification of the eRGMB-NEO1_{FN56} complex, the binding affinities of eRGMB alone ($K_d = 416$ nM) and the eRGMB-NEO1_{FN56} complex ($K_d = 407$ nM) to the BMP2 ligand were compared (Fig. 5.6b,c). The presence of NEO1_{FN56} in the eRGMB-NEO1_{FN56} complex did not appear to significantly increase the affinity of eRGMB for BMP2 binding, suggesting that the NEO1_{FN5}-BMP2 interface observed in the crystal structure does not function *in vivo* to enhance the RGM-BMP interaction. The effect of full-length NEO1 on BMP signalling was subsequently tested in a BRE-Luc BMP luciferase reporter assay. LLC-PK1 cells were co-transfected with control plasmid, or plasmids for the expression of either full-length RGMB or NEO1 and stimulated with 6 nM BMP2. As expected, RGMB enhanced the BMP2 response by approximately three-fold ($n=16$, $P<0.0001$), however, NEO1 induced a massive increase in BMP signalling, by approximately ten-fold over the control ($n=16$, $P<0.0001$). It seems clear then that NEO1 has an activating effect on BMP signalling, however, due to time constraints we were unable to fully investigate this effect.

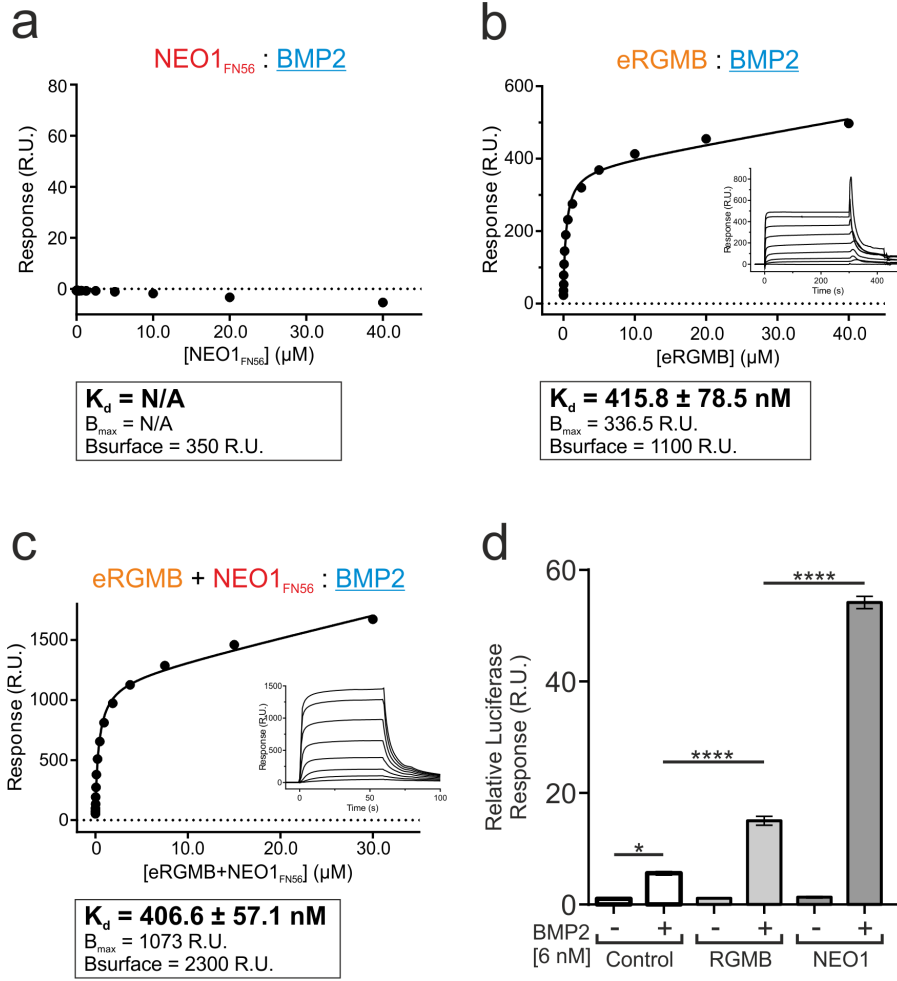


Figure 5.6: Probing the role of NEO1 in BMP signalling. (a-c) SPR equilibrium analysis of BMP2 binding to (a) NEO1_{FN56}, (b) eRGMB and (c) eRGMB-NEO1_{FN56}. Graphs show a plot of the equilibrium binding response (R.U. = response units) against concentration of the analytes used (μM). Binding constants (K_d) are given as the mean $\pm$ the standard error of the mean (n=2 technical repeats). Fits are shown as lines. Sensorgrams are shown as plots of response (R.U.) against time (s). Binding parameters K_d , B_{max} and surface response units (B_{surface}) are indicated below the graphs. (d) BRE-Luc BMP2 luciferase response assay with co-transfected GPI-anchored RGMB and full-length NEO1; data are from two independent experiments. Error bars, s.e.m. (n = 16 cell cultures in each group), **** $P < 0.0001$, * $P = 0.0192$ by one-way ANOVA with a Tukey multiple comparison test.

5.3.2 The solution structure of the BMP2-eRGMB-NEO1_{FN56} complex

To investigate the oligomeric arrangement of the BMP2-eRGMB-NEO1_{FN56} ternary complex in solution and to compare the complex architecture with that observed in the crystal structure a series of SAXS experiments were conducted. In preparation for analysis of the SAXS data, complete models of the individual protein components were prepared. For all of the complexes missing loops, unresolved residues at the N- and C-termini and cloning tags were added in an extended conformation using the program Modeller[263] and one N-linked glycan was built on each of NEO1_{FN5} and RGMB_{ND} domains. For RGMB, a high resolution crystal structure for the C-terminal domain has still not been obtained; for example, this domain was not resolved in the previously solved eRGMB-NEO1_{FN56} complex crystal structure, and therefore, a structural model for this domain was calculated using the program ROSETTA3.5[262]. The secondary structure of this domain was predicted to be α -helical with one disulphide bond between C358 and C372, and so the model was constrained within these parameters. The lowest energy solution that also demonstrated a compact three-helix structure and a hydrophobic core was selected (Fig. 5.7b,c, right panels).

Initially, SAXS analysis was conducted for each of the individual components and the binary complexes. Towards this, NEO1_{FN56}, eRGMB, RGMB_{ND}-BMP2 and eRGMB-NEO1_{FN56} were SEC-purified and SAXS data were collected at the ESRF beamline BM29 for each sample over a range of concentrations to test for concentration-dependent aggregation. Final scattering curves were generated by merging the low angle part of the best low concentration dataset and the high angle part of the best high concentration dataset for each sample; corresponding to data collected at the following concentrations: NEO1_{FN56}: 1.06 and 4.94 mg/ml; eRGMB: 0.96 and 5.48 mg/ml; eRGMB-NEO1_{FN56}: 0.51 and 1.01 mg/ml

(Fig. 5.7, top left panels, black curves). These scattering curves were used to derive the pair-distribution function $P(r)$, and the Guinier region (Fig. 5.7, black curves). This analysis allowed several important parameters describing the shape and size of the scattering particles in the sample to be derived.

From the pair distribution function $P(r)$, the maximum linear dimension of the scattering particle ($D_{\max}$) was calculated. The radius of gyration (R_g) and the point at which the scattering curve intersects the y-axis ($I(0)$) were determined using the Guinier plot. In all cases a small concentration-dependent aggregation effect was observed for the high concentration samples (as indicated by a lack of linearity in the Guinier region) (data not shown). However, due to recent advances in SAXS beam-line technology, even the datasets recorded for samples with concentrations lower than 1 mg/ml were of high quality with good signal-to-noise levels. For these low concentration samples a linear plot of the Guinier region was obtained, indicative of a homogenous, monodisperse sample lacking long-range interactions in solution, which allowed accurate determination of $I(0)$ and R_g . Interestingly, the data collected for the RGMB_{ND} -BMP2 complex showed high levels of aggregation, most likely due to the BMP2 ligand which has been well characterised as a sticky protein at neutral pH, and this made analysis of the data for this complex impossible (data not shown).

In addition, a reference-free assessment of the molecular weight of the scattering particle was calculated for each sample using the volume of correlation (V_c) metric. These determined masses compared well to the expected molecular weights. The experimental molecular weight measurement for $\text{NEO1}_{\text{FN56}}$ was 24.0 kDa, which was very similar to the calculated mass of glycosylated $\text{NEO1}_{\text{FN56}}$ (25.54 kDa) (Fig. 5.7a); the measurement for eRGMB was 47.2 kDa which again compared favourably to the calculated mass of 42.76 kDa (Fig. 5.7b); and the value for the eRGMB- $\text{NEO1}_{\text{FN56}}$ complex was 68.3 kDa, which was in good agreement with the calculated molecular weight for this complex (68.3 kDa) (Fig. 5.7c).

This analysis of the molecular weights of the scattering particles confirmed the expected solution stoichiometry of each protein as monomeric and the eRGMB-NEO1_{FN56} complex as 1:1.

To analyse the domain conformations sampled by each structure in solution, given the structural models based on the crystal structures of each protein or protein complex, a previously described minimal ensemble search approach[266, 259] was utilised. In this method a coarse-grained molecular dynamics simulation was performed using each model protein structure, as implemented in the AllosMod[264] program, and a large pool of models was generated representing snapshots from throughout the simulation, capturing the domain movements. The program FoXS[265] was used to calculate theoretical scattering curves for each model and the program MES[266, 259] was used to select the minimal ensemble of models that most accurately fits the experimental data. This approach attempts to limit overfitting as much as possible. The fit of the calculated scattering curve (Fig. 5.7, red curves) for this ensemble of models was assessed through a plot of the residuals across the scattering range and a χ^2 value describing the overall goodness of fit.

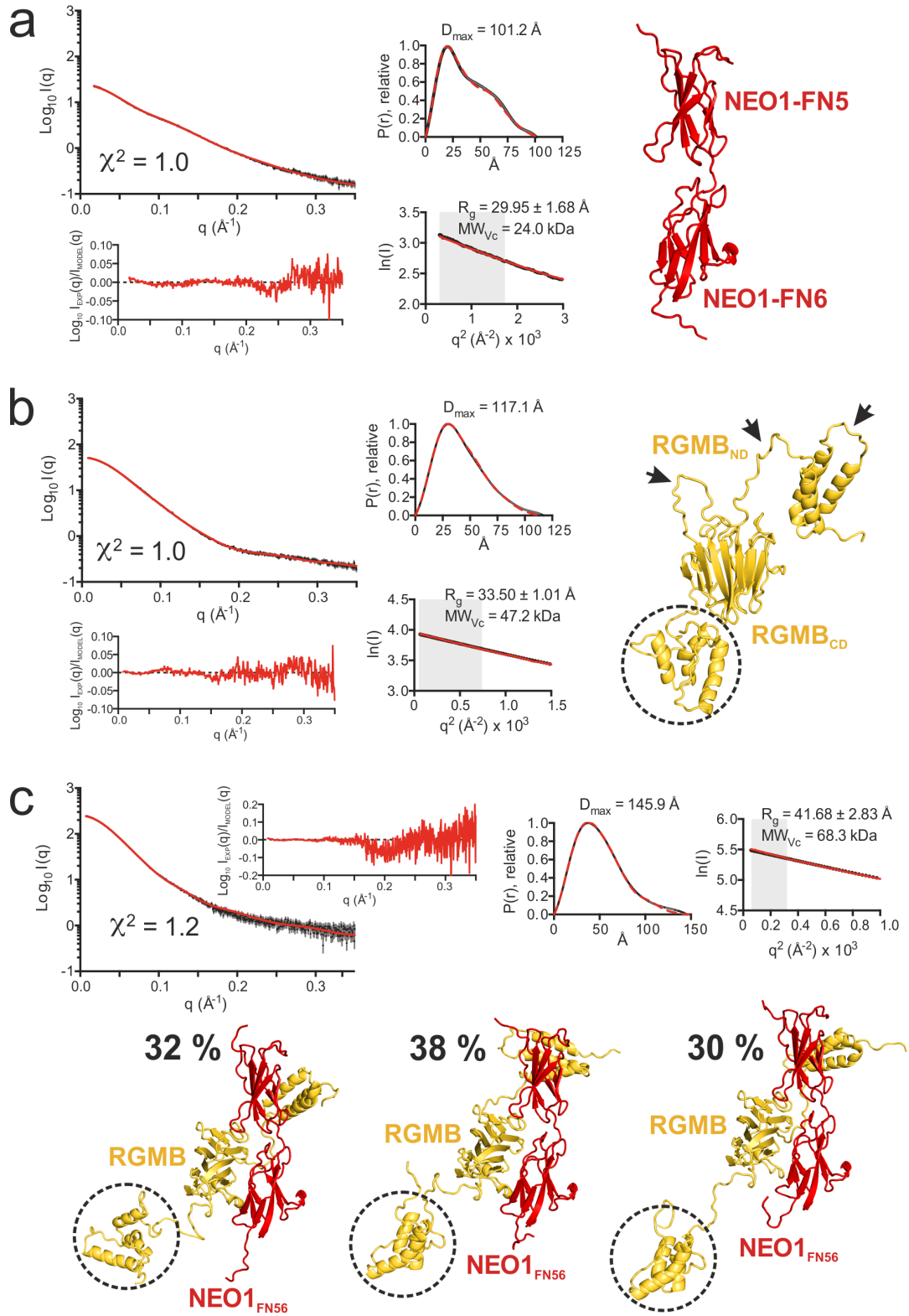


Figure 5.7: SAXS analysis of NEO1_{FN56}, eRGMB and the eRGMB-NEO1_{FN56} complex. SAXS analysis of (a) NEO1_{FN56}, (b) eRGMB and (c) eRGMB-NEO1_{FN56} after isolation by SEC. Experimental scattering curves (black) and calculated scattering patterns (red) are shown to a maximal momentum transfer of $q = 0.35 \text{ \AA}^{-1}$ in the upper left panels. The fitting residuals of the experimental scattering curves and calculated scattering patterns are displayed. The experimental (black) and calculated (red) normalised pair distribution ($P(r)$) functions, yielding the maximum intra-particle diameter ($D_{\max}$); and the experimental (black) and calculated (red) Guinier regions, yielding the radius of gyration (R_g), and the molecular weight derived from the volume of correlation metric V_c (MWV_c) are shown. The shaded area indicates the range of fitting for R_g analysis ($RG\bullet S \leq 1.3$). The fit of the calculated scattering patterns (χ^2) are annotated. The models and their fractional contribution to the scattering pattern fit are displayed in cartoon representation. NEO1_{FN56} is red, eRGMB is gold, black arrows indicate modelled linkers, a black dotted circle highlights the model for the RGMB C-terminal domain.

For NEO1_{FN56}, one model could accurately describe the experimental scattering data ($\chi^2 = 1.0$) (Fig. 5.7a). Similarly for eRGMB, one model was sufficient to accurately describe the experimental scattering data ($\chi^2 = 1.0$), suggesting that the *ab initio* model constructed for the C-terminal domain was a reasonable assumption for this structure (Fig. 5.7b). The scattering data for the eRGMB-NEO1_{FN56} complex required an ensemble of three models for accurate curve fitting ($\chi^2 = 1.2$), and an analysis of these models suggests that there is flexibility in the eRGMB linker regions in this complex (Fig. 5.7c). Overall these SAXS data allowed the stoichiometry of the sub-complexes and individual protein components to be confirmed, and allowed complex formation to be followed demonstrating that the molecular weight of the complex was consistent with the sum of the individual masses. Furthermore, this analysis validated the structural models produced for the individual protein components and confirmed the effectiveness of the minimal ensemble search method for analysing the solution states of these proteins and their complexes. This initial analysis laid the foundation for the SAXS analysis of the ternary BMP2-eRGMB-NEO1_{FN56} complex.

The ternary BMP2-eRGMB-NEO1_{FN56} complex was SEC-purified and SAXS

data were collected at the ESRF beamline BM29 at two concentrations: 0.68 and 1.14 mg/ml. Once again some aggregation was observed for the high concentration sample but this was markedly decreased in the low concentration sample. The final plot of the Guinier region for this sample at 0.68 mg/ml was linear, indicating a monodisperse sample with few long-range solution interactions. It is likely that in the ternary complex, surrounded by binding partners the stickiness of the BMP2 ligand is greatly reduced compared to the BMP2-RGMB_{ND} complex where large portions of the ligand remain exposed, and this allowed good quality data for the ternary complex to be collected. The experimental estimate of the molecular weight for the BMP2-eRGMB-NEO1_{FN56} complex was 168.0 kDa, which is in good agreement with the calculated molecular weight for this complex of 162.4 kDa. This analysis suggests that the solution stoichiometry of the ternary complex is 2:2:2, which replicates that observed in the crystal structure. Regarding the arrangement of the ternary complex in solution, a comparison of the theoretical scattering curve for the crystal structure and the experimental scattering curve revealed that the fit was very poor ($\chi^2 = 28.1$) (Fig. 5.8a). However, incorporating the complete models for each protein, including missing loops, termini, sugars and the eRGMB C-terminal domain model; and following implementation of the minimal ensemble search protocol as described above, the experimental scattering curve could be accurately described by two models of the ternary complex ($\chi^2 = 1.9$) (Fig. 5.8b). The models of the ternary complex that fit the scattering data show marked similarity to the crystal structure, with the overall architecture largely conserved, although there is relative movement of the individual domains that appears to be primarily mediated through the flexible linkers connecting RGMB_{ND} and RGMCD (Fig. 5.8). The BMP2-BMP2, eRGMB-NEO1_{FN56} and RGMB_{ND}-BMP2 interfaces seem to be preserved in solution, however, one copy of the NEO1_{FN5} domain dissociated from the BMP2 dimer over the course of the molecular dynamics simulation (indicated by an arrow in Fig. 5.8b). This supports the conclusion derived from the interface

analysis that the NEO1_{FN5}-BMP2 interface is not significant and is likely not to be physiologically important.

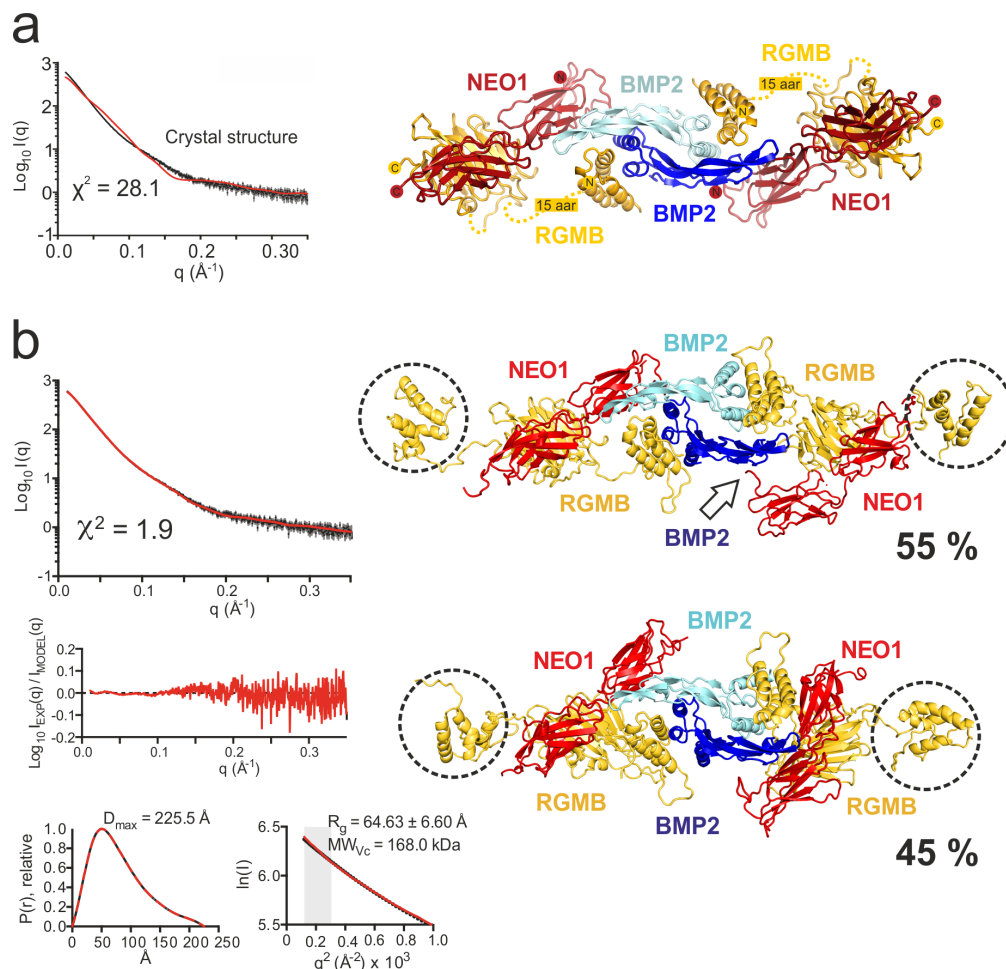


Figure 5.8: SAXS analysis of the BMP2-eRGMB-NEO1_{FN56} complex. (a) Fitting of the calculated scattering pattern of the BMP2-eRGMB-NEO1_{FN56} crystal structure (red) to the experimental scattering data (black) (left panel). The goodness of fit (χ^2) is indicated. Cartoon representation of the crystal structure (right panel). NEO1_{FN56} is red, eRGMB is gold, BMP2 is blue/cyan. (b) Fitting of minimal ensemble models of BMP2-eRGMB-NEO1_{FN56} to the SAXS data. The experimental (black) and calculated (red) scattering curves are shown to a maximal momentum transfer of $q = 0.35 \text{ \AA}^{-1}$ and the fitting residuals are displayed below. The experimental (black) and calculated (red) normalised pair distribution ($P(r)$) function (lower left panel) and Guinier region (lower right panel) are displayed. The shaded area indicates the range of fitting for R_g analysis ($RG \bullet S \leq 1.3$). The goodness of fit (χ^2), derived maximum intra-particle diameter ($D_{\max}$), radius of gyration (R_g), and molecular weight derived from the volume of correlation metric V_c (MW_{Vc}) are annotated. The models, and their fractional contribution to fitting the scattering data, are displayed in cartoon representation, colouring as in (a). Black dotted circle highlights the RGMB C-terminal domain model and an arrow indicates dissociation of NEO1_{FN5} from BMP2.

5.3.3 BMP2 mediates clustering of NEO1 on the cell membrane

The SAXS data demonstrated that the BMP2-eRGMB-NEO1_{FN56} ternary complex adopts a 2:2:2 stoichiometry in solution with an architecture similar to that observed in the crystal structure; however, the function of this complex *in vivo* is not yet clear. An analysis of the crystal structures intriguingly showed that the architecture of the ternary complex facilitates the incorporation of the previously determined structure of the active eRGMB-NEO1_{FN56} signalling hub (Fig. 5.9). A superposition of these two complexes results in two copies of the 2:2 stoichiometry eRGMB-NEO1_{FN56} complex linked by the BMP2 dimer. This binding mode can be further amplified, demonstrating how BMP2 could potentially link together RGM-NEO1 complexes in rafts along the cell membrane (Fig. 5.9).

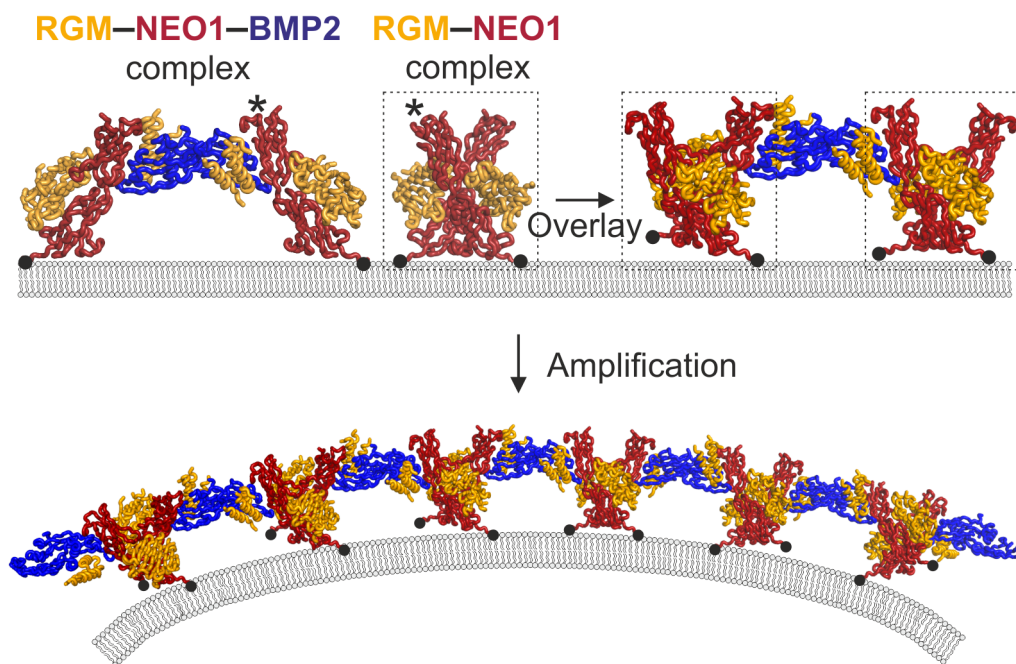


Figure 5.9: A model of RGM-mediated clustering based on crystal-structure analysis. Reprinted by permission from Macmillan Publishers Ltd: *Nature Structural and Molecular Biology*, [320] copyright (2015). Superposition of the ternary BMP2-eRGMB-NEO1_{FN56} and the active signalling conformation of the eRGMB-NEO1_{FN56} complex (PDB ID: 4BQ6[240]) using NEO1_{FN56} (marked with an asterisk) as the template. Amplification of this arrangement results in a continuous array in which RGMB bridges the dimers of BMP2 and NEO1. The C-termini of NEO1 and RGMB are marked by black dots; the membrane is indicated by a cartoon phospholipid bilayer (grey).

TIRF-dSTORM imaging was then used to test this clustering model; this technique allows the x-y localisation of single fluorescent molecules, specifically within the cell membrane, to be recorded to a precision of around 30 nm. Towards this, COS7 cells were co-transfected with GPI-anchored RGMB and full-length NEO1 fused C-terminally to an mVenus fluorescent tag. By imaging using a fluorescence microscope in TIRF mode, only NEO1-mVenus molecules localised to the plasma membrane were detected and this reduced the background signal (Fig. 5.10a). A high-powered laser was used to transiently transfer the majority of the mVenus fluorophores into a recoverable dark state, resulting in only a small number of photons being detected at any one time. The PSF of these fluorescent emission events could be interpreted as 2D Gaussians, allowing the fluorophores to be more precisely localised to the peak of the distribution. Over a data collection period of approximately 20 minutes, the precise localisation of many individual NEO1-mVenus molecules could be determined and small regions of interest were selected for analysis (Fig. 5.10b). The NEO1 receptors appeared to cluster in the membrane (Fig. 5.10b, right panels) and the level of clustering was analysed using Ripley's K -function (Fig. 5.10c,d), with the peak of this curve, also known as the $L(t)$ max value, indicating the level of clustering[268, 269]. TIRF-dSTORM experiments in both live and fixed cells were performed to analyse the change in the clustering levels of NEO1-mVenus in the presence of RGMB after the addition of BMP2 (Fig. 5.10 and Fig. 5.11). For the fixed-cell analysis, samples were taken at time points after BMP2 addition and the clustering levels in each one analysed; whereas for the live-cell imaging, BMP2 was added and the clustering level analysed used a time-gated approach[268, 275].

A representative example of two live cells is shown, demonstrating the clustering of NEO1-mVenus over time following the addition of either BMP2 or buffer (Fig. 5.10). This example clearly demonstrates that upon addition of BMP2 the clustering level of NEO1-mVenus increases. Plots of Ripley's K -function

for each cell over the consecutive time windows show that the NEO1 clustering appears to peak 6-13 minutes after BMP2 addition and then to decrease after 10-17 minutes (Fig. 5.10c); whereas in the control sample, the clustering level does not appear to change (Fig. 5.10d). A plot showing the maxima of Ripley's K -function, $L(t)_{\max}$, for the BMP2-treated and control cells during each time window shows a clear peak in the NEO1-mVenus clustering in the BMP2-treated sample that is not visible in the control (Fig. 5.10e). The first $L(t)_{\max}$ value is recorded approximately 10 minutes after BMP2 addition, this time delay is necessary as the number of photon counts must reach a threshold before the first time window can be allocated for analysis.

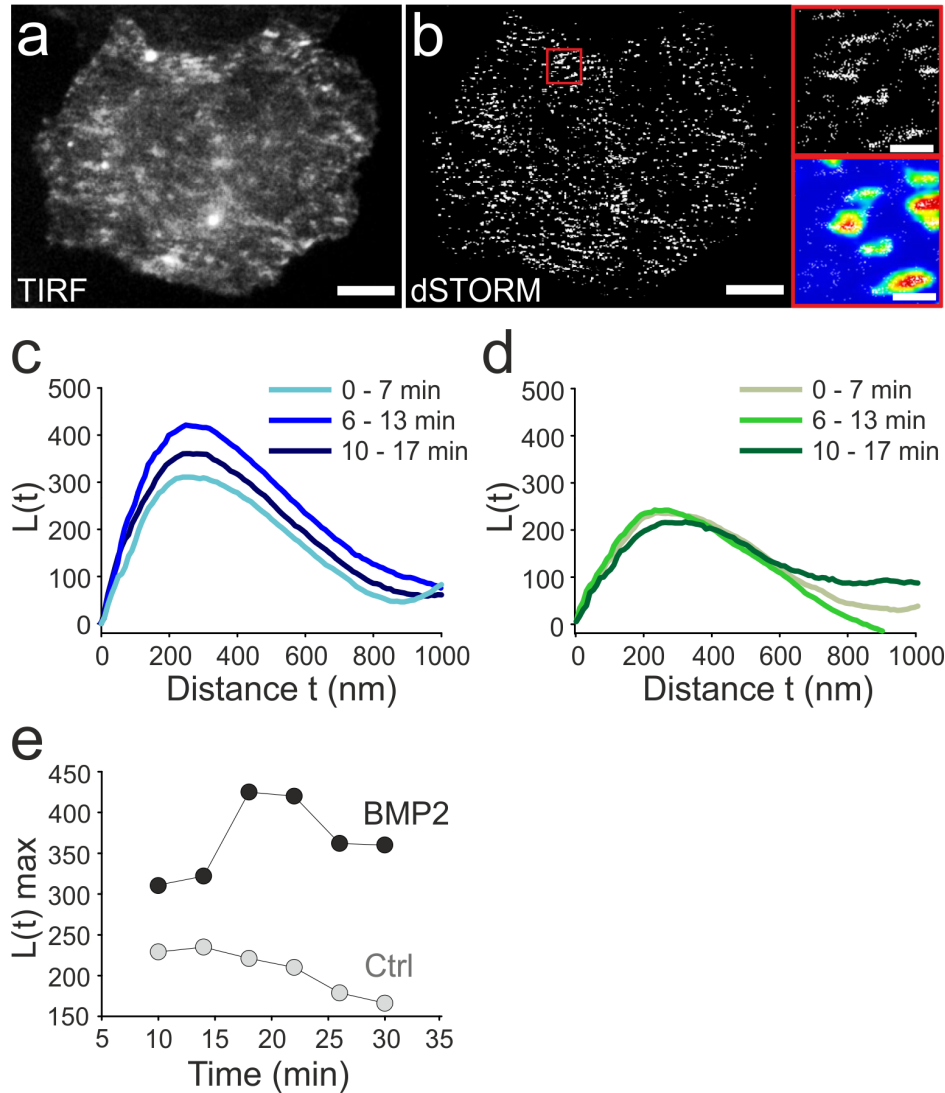


Figure 5.10: A representative example of the live-cell analysis of BMP2-induced NEO1-mVenus clustering. Reprinted by permission from Macmillan Publishers Ltd: [Nature Structural and Molecular Biology](#), [320] copyright (2015). (a-b) Live-cell TIRF (a) and live-cell TIRF-dSTORM (b) images of a COS7 cell co-expressing NEO1-mVenus and RGMb, treated with BMP2. Scale bars, 10 μm . Right panels of (b) show cluster maps generated from a $5 \times 5 \mu\text{m}$ region of interest (ROI) (red square in (b)), with scale bars of 1.5 μm . Bottom right panel shows a cluster heat map coloured from blue (low clustering) to red (high clustering). (c-d) Ripley's K -function ($L(t)$), calculated from an ROI within a single cell treated with either BMP2 (c) or buffer control (d). Time-gated window ~ 7 min (2,000 frames). (e) Plot of the maxima of Ripley's K -function ($L(t) \max$) calculated from an ROI within a single cell over the indicated time windows after the addition of either BMP2 or buffer control (Ctrl). BMP2-treated (black dots) and control (white dots) cells are labelled.

These live-cell TIRF-dSTORM experiments were performed for 8 cells treated with BMP2 and 6 cells treated with a buffer control. A plot of the mean $L(t)_{\max}$ for each time window shows increased clustering of NEO1-mVenus in the BMP2-treated cells compared to the control samples (Fig. 5.11a). The largest significant difference in mean $L(t)_{\max}$ value between the control samples and the BMP2-treated samples occurred at approximately 18 minutes, suggesting that at this point the BMP2-induced clustering of NEO1-mVenus was maximal. These experiments were also performed in fixed cells where samples were taken 0, 0.5, 1, 5 and 20 minutes after addition of BMP2 or buffer, and the level of NEO1-mVenus clustering was analysed for 6-10 cells at each time point (Fig. 5.11b,c). The plots of mean $L(t)_{\max}$ for the BMP2-treated (Fig. 5.11b) and control samples (Fig. 5.11c) at each time-point show that the clustering level of NEO1-mVenus was highest 5 minutes after addition of BMP2; at this point the mean $L(t)_{\max}$ value was significantly different for that observed in the control samples ($n=10$, $P = 0.02$) (Fig. 5.11b). This analysis suggests that addition of BMP2 induced clustering of NEO1-mVenus, which reached a maximal level approximately 5 minutes after addition of the ligand (Fig. 5.11b). In comparison the addition of buffer did not significantly alter the level of NEO1-mVenus clustering (Fig. 5.11c). These imaging data appear to support the clustering model, suggesting that BMP2 induces RGM-mediated clustering of NEO1 receptors in a time dependent manner.

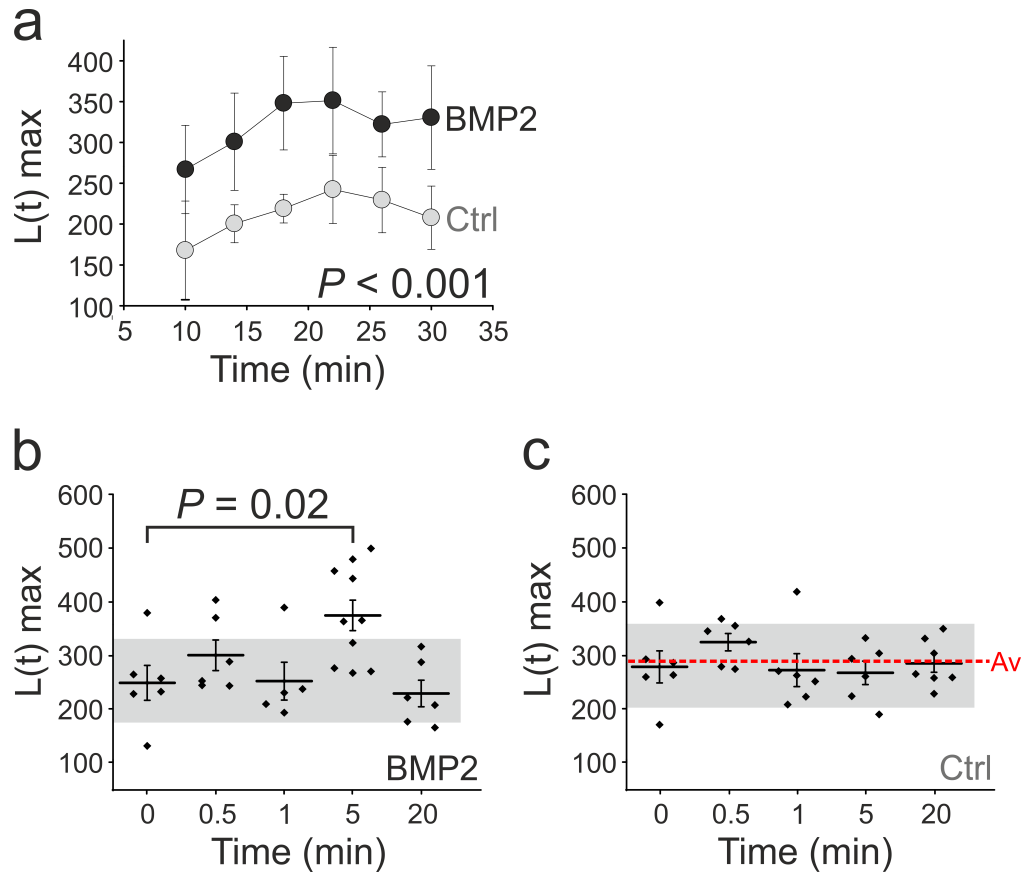


Figure 5.11: Live- and fixed-cell analysis of BMP2-induced NEO1 clustering. (a) Mean $L(t)_{\max}$ calculated for all analysed live cells treated with BMP2 (black dots, $n = 8$ cells) or buffer control (white dots, $n = 6$ cells). BMP2 addition was at time 0 min. Error bars, s.d. $P < 0.001$ by two-tailed Student's t test. (b-c) Plots of $L(t)_{\max}$ of regions of interest from populations of cells fixed at different time points (in minutes (min)) after treatment with (b) BMP2 or (c) buffer control. Each point represents one cell. The horizontal lines represent the mean $L(t)_{\max}$. Error bars, s.e.m. The dotted red line highlights the average $L(t)_{\max}$ (255) and the grey rectangles the s.d. (± 55) for the control. $P = 0.02$ at $t = 5$ min in the BMP2-treated cells, by two-tailed Student's t test. For the BMP2-treated cells, $n = 6, 6, 5, 10$ and 6 cells, from left to right; for the control-treated cells, $n = 6, 6, 6, 6$ and 7 cells, from left to right.

5.4 Discussion

5.4.1 RGM mediates the interaction between BMP and NEO1 signalling

Although there are many studies showing that the interdependence of BMP and NEO1 signalling is mediated through RGMs this was still a point for debate. A recent publication showed that NEO1 could bind directly to BMP ligands and that NEO1 functioned as a BMP receptor[321]. Given the findings outlined in this chapter, this now seems unlikely. In the ternary BMP2-eRGMB-NEO1_{FN56} crystal structure a contact between BMP2 and the NEO1-FN5 domain was observed, however an interface analysis suggested that this was unlikely to be important for complex assembly (Fig. 5.5). Furthermore, SPR equilibrium binding experiments showed no direct interaction between BMP2 and NEO1_{FN56}, and NEO1_{FN56} did not appear to enhance binding of eRGMB to BMP2 (Fig. 5.6). In addition, the BMP2-NEO1_{FN5} interaction dissociated over the course of a molecular dynamics simulation which generated one of the models of the ternary complex in the ensemble that fitted the experimental SAXS data, suggesting that this interaction is not thermodynamically favoured (Fig. 5.8). Overall these data suggest that *in vivo* BMP2 and NEO1 are unlikely to associate directly, thereby supporting the hypothesis that RGM forms a functional bridge between the two pathways. Although, whether RGM-activated BMP signalling can occur in the absence of NEO1, and *vice versa*, whether RGM signalling through NEO1 can be fully active in the absence of BMPs remains to be seen and different studies have come to different conclusions regarding this[160, 192, 191]. It would be interesting to test whether RGM mutants that are unable to bind to NEO1, for example the A186R^{RGMB} mutant, can still activate BMP signalling, perhaps using the BRE-Luc BMP luciferase reporter assay. Similarly, RGM mutants that are un-

able to bind to BMP, for example the H106R^{RGMB} mutation, could be tested for their ability to trigger growth cone collapse in neurite outgrowth inhibition assays.

5.4.2 BMP-induced clustering of NEO1 may be physiologically important

The crystal structure of the ternary BMP2-eRGMB-NEO1_{FN56} complex showed that RGM can simultaneously bind to both NEO1 and BMP2 and a superposition of this structure with that of the active signalling conformation of the eRGMB-NEO1_{FN56} complex suggested that RGM might cluster dimers of NEO1 and BMP2 in arrays on the membrane (Fig. 5.9). TIRF-dSTORM microscopy experiments supported this hypothesis, showing that addition of BMP2 induced clustering of NEO1 receptors in the presence of RGMB (Fig. 5.10 and Fig. 5.11). However, the physiological importance of this clustering is not yet clear. Several studies have indicated that the targeting of key players in NEO1 and BMP signalling to membrane microdomains is important for activation of these pathways[198, 197]. It would be interesting to link this clustering to activation of either the BMP2 or NEO1 signalling pathway. This could potentially be tested by disrupting the dimeric RGM-NEO1 interaction; as described in Chapter 3 this could be done using the P206N^{RGMB} mutation that disrupts the formation of the site-2 RGMB-NEO1 interface that is important for 2:2 stoichiometry complex formation (Fig. 3.12). This mutant would prevent RGMB-NEO1 dimers forming and thereby disrupt the clustering without disrupting the site-1 RGMB-NEO1 interaction. This mutant could be used to assay for its effect on BMP signalling, perhaps using the BRE-Luc BMP luciferase assay. However, as shown above (Fig. 5.6d) transfection of NEO1 into LLC-PK1 cells had an unexpected effect in this assay; NEO1 enhanced BMP signalling by approximately ten-fold,

compared to transfection of RGMB which only enhanced the signalling by three-fold. The reason for this difference is not clear, for example we do not know what the endogenous expression levels are of RGMB or NEO1 in this cell type, and until the two effects can be de-convoluted it will be difficult to test RGMB or NEO1 mutants in this assay.

5.5 Summary

This chapter describes the structure solution of the BMP2-eRGMB-NEO1_{FN56} ternary complex and shows for the first time that RGM acts as a physical link bridging the two fundamental signalling pathways of BMP and NEO1 signalling. This provides a structural rationale for the evidence that NEO1 and BMP signalling are functionally linked. Crystal structure interface analysis, SPR experiments and molecular dynamics simulations have also supported the requirement of RGMs in linking BMP and NEO1 signalling, showing that NEO1 and BMP2 are unlikely to associate independently of RGMs. Furthermore, super-resolution fluorescence microscopy experiments have also shown that RGMs mediate BMP2 dimer-induced clustering of NEO1 receptors in the membrane to form ordered arrays that may be important for NEO1 and BMP signalling.

Chapter 6

Discussion

6.1 The modular domain architecture of RGMs

Prior to the X-ray crystallography experiments conducted in our laboratory there was no structural information available for the RGM family. RGMs were known to bind specifically to the NEO1 receptor and also to BMP ligands. The first insight into the structure of the RGM proteins came from the eRGMB-NEO1_{FN56} complex structure[240] (Fig. 3.2). In this structure, despite using the full-length ectodomain of RGMB in the crystallisation experiments, only the middle domain of RGMB could be resolved in the electron density map of the complex structure. The crystal structure revealed that the RGMB middle domain adopts the form of a tightly packed β -sandwich followed by four short α -helices, with the arrangement stabilised by three disulphide bonds (Fig. 3.3). From sequence analysis a partial von Willebrand factor type D (vWfD) domain had previously been recognised within this region of the protein[36, 50]. The vWfD domain containing superfamily of proteins, identified through sequence similarity, spans functions from cell adhesion (e.g. mucins[323, 324]), to axon growth (e.g. SCO-Spondin[325]) to regulators of BMP signalling (e.g. Kielin[326] and BMPER[326]). There is no structural information for this domain and it shares no structural homology with any other protein domain. Since the RGMB middle domain is potentially a modified vWfD domain this crystal structure represents the first insight into the potential architecture of this domain.

As outlined in Chapter 3, the NEO1 binding site was localised to one face of the β -sandwich of the RGMB middle domain, and in the structure the N- and C-termini of RGMB project away from the domain into solvent channels in the crystal implying that the N- and C-terminal regions of the protein function as independent structural units (Fig. 3.2 and 3.3). Following this, as presented in Chapter 4, structures of RGM_{ND} in complex with the BMP2 ligand were solved (Fig. 4.9) and finally the structure of the ternary BMP2-eRGMB-NEO1_{FN56}

complex revealed that RGMs can simultaneously bind to both NEO1 and BMP2. In this structure the N-terminal and middle domains of RGMB are joined by a flexible amino acid linker (Fig. 5.4) and in the molecular dynamics simulations outlined in Chapter 5 the two domains were shown to move relative to each other, further suggesting that they function independently (Fig. 5.8). The final domain of RGMs for which there is still no structural information is the C-terminal domain. As described in Chapter 5, a ROSETTA model of the RGM C-terminal domain was constructed in which the domain was constrained to be α -helical with a disulphide bond between C358 and C372. The overall structural model of the eRGMB protein, including the C-terminal domain, was a good fit to the experimental SAXS curve for this protein (Fig. 5.7b). However, due to the low resolution of SAXS as a technique, this is no guarantee that the modelled structure of the C-terminal domain is correct. High-resolution structural data for this domain, obtained by X-ray crystallography or NMR, is therefore still required to obtain a complete structural picture for the RGM family.

In combination with its unknown structure, the function of this C-terminal domain is also unknown. An RGM $_{\Delta CD}$ construct does not express in mammalian cells (data not shown), in contrast to the RGM $_{\Delta ND}$ construct, which implies that the C-terminal domain has an important role in the folding of the middle domain. It is also plausible, since both the N-terminal and middle domains of RGM function as discrete units with different interacting partners that the C-terminal domain may also have an as yet unidentified protein interaction partner (Fig. 6.1). Another potential site of protein interaction on RGMs is the face of the β -sandwich opposing the NEO1 binding site; this surface has an attractive groove running down the centre of the β -sandwich, which contains a high density of hydrophobic residues, such that it has acquired the nickname the “phenylalanine field” (Fig. 6.1, green surface), it therefore appears to be a likely candidate region for a protein binding site. Another possibility is that this region is the

binding site for glycosaminoglycans (GAGs), complex polysaccharides that are found on both the cell surface and in the extracellular matrix. Another repulsive axon guidance system, the Slit-Robo signalling pathway was shown to be dependent on the abundant GAG, Heparan Sulphate[327, 328, 329]; and GAGs have also been shown to be important in growth factor signalling, for example the FGF[330] and TGF β [331] pathways. Heparin-affinity chromatography demonstrated that eRGMB binds to heparin (data not shown), suggesting that soluble forms of RGMs, in a physiological context, would likely associate with GAGs such as Heparan Sulphate. The binding site of these GAGs on RGMs has not yet been determined and it would be interesting to conduct a crystallisation experiment of an eRGMB-heparin complex.

RGMs are known to fulfil a plethora of functions through mammalian development into the adult and the work presented in this thesis has characterised them as a node for the BMP and NEO1 signalling networks. It is certainly therefore conceivable that other RGM-interacting proteins have yet to be characterised, for example it was recently suggested that programmed death ligand 2 (PD-L2), an inhibitor of T-cell activation in the immune system, is a novel binding partner of RGMB[332]. This is consistent with the known roles of RGMs in the immune system, which range from activation of T-cells[79] to the control of leukocyte chemotaxis[78] and the inhibition of interleukin expression[88].

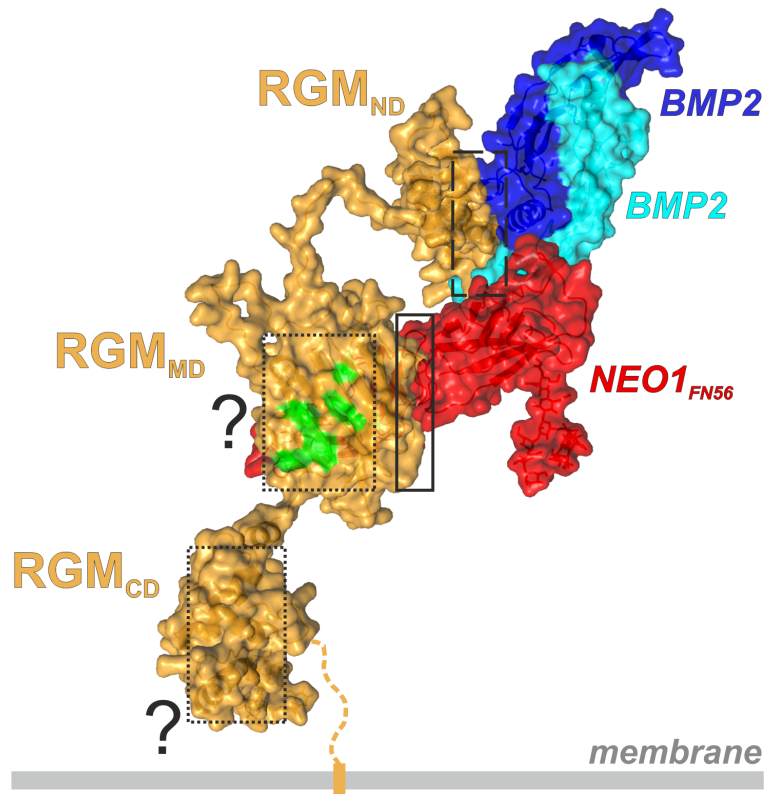


Figure 6.1: The modular architecture of RGM and its interaction sites. A transparent solvent-accessible surface and cartoon representation of RGM (gold) with flexible linkers and the C-terminal domain modelled. Transparent surface and cartoon representations of BMP2 (blue and cyan) and NEO1_{FN56} (red, N-linked glycan shown in stick representation) binding to RGM are shown. The RGM N-terminal (RGM_{ND}), middle (RGM_{MD}) and C-terminal (RGM_{CD}) domains are labelled. The confirmed binding sites are outlined with black boxes; potential interaction sites are outlined with dotted line boxes and labelled with question marks. The solvent exposed phenylalanine residues in the RGM middle domain are shown as green surface.

6.2 Insights into Netrin-1 signalling from the RGM-NEO1 structure

The structural analysis of the eRGM-NEO1_{FN56} complex structure (Fig. 3.8 and 3.9), in combination with AUC experiments (Fig. 3.12) demonstrated that eRGM and NEO1_{FN56} form a 2:2 complex, which was shown using neurite outgrowth inhibition assays (Fig. 3.15) to be important for signalling. RGM

acts as a molecular staple bringing together and fixing the orientation of two NEO1 receptors to activate downstream signalling. NEO1 is not only a receptor for RGMs but also for the attractive guidance cue Netrin-1[143]. A recent crystal structure of the NEO1_{FN45}-Netrin-1 complex showed that Netrin-1 dimers bind in an X-shape, connecting two NEO1 receptors[147]. Despite RGM and Netrin-1 both bringing together NEO1 dimers the architecture of the two complexes is very different with RGM inducing a signalling competent form of the NEO1 receptors in which the cytoplasmic domains would likely be closely associated (< 60 Å apart); as opposed to Netrin-1 which imposes a substantial distance between the two receptors (~ 105 Å) that might inhibit NEO1 signalling (Fig. 6.2). Netrin-1 binds at the FN4 and FN5 domains of NEO1 whereas RGM binds at the FN5 and FN6 domains, and the two binding sites do not overlap. It is possible therefore that RGM and Netrin-1 bind NEO1 at the same time but the physiological effect of this remains to be seen (Fig. 6.2).

The primary receptor for Netrin-1 is the NEO1 homologue DCC and binding of Netrin-1 to DCC has been shown to mediate axon outgrowth[143]. DCC and NEO1 are very similar; they have an identical extracellular domain structure and share a sequence identity of 50 %. As expected therefore, the FN4-5 domains of NEO1 and DCC have similar binding affinities for Netrin-1 ($K_d = 24$ nM and $K_d = 23$ nM respectively[147]); however, interestingly this is not the case for RGM binding. The binding affinity of DCC for RGM is about one thousandth of that for NEO1[240]. As described in Chapter 3, this is likely due to the L3-loop in the NEO1 FN5 domain, which is very poorly conserved in DCC, and therefore acts as a specificity determinant for RGM binding[240]. NEO1 has therefore evolved a secondary level of signalling through RGMs that is not shared by DCC, suggesting that control of axon guidance by NEO1 is more complex as it is able to receive signalling inputs from two different ligands.

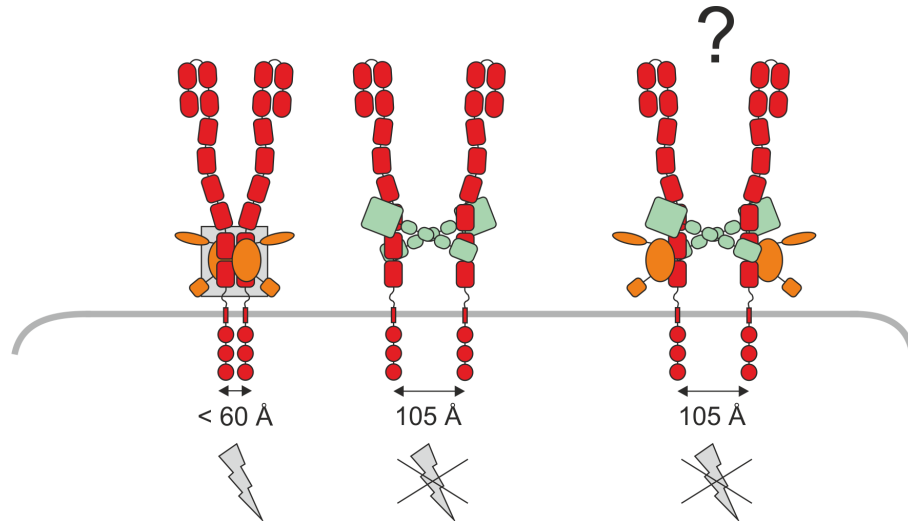


Figure 6.2: The different binding modes of RGM and Netrin-1 to NEO1. Schematic showing the binding modes of RGM and Netrin-1 to NEO1 as inferred from crystal structures. Potentially RGM and Netrin-1 can bind NEO1 simultaneously. NEO1 = red, RGM = orange, Netrin-1 = green, grey lightning bolts indicate NEO1 signalling, the estimated distance between the NEO1 cytoplasmic domains is indicated.

6.3 The RGM-BMP complex may preclude BMP receptor binding

In Chapter 4 the structure of the RGM N-terminal domain in complex with the BMP2 dimer was presented (Fig. 4.9). All RGM family members are known activators of BMP signalling[102, 83, 89] although the molecular mechanism by which this activation occurs is not known. In Chapter 4 an endocytosis-linked mechanism was suggested for RGM-activated BMP signalling (Fig. 4.24), however other potential mechanisms have been outlined in several publications, some of which are compatible with the structural data presented in this thesis. An early model proposed that RGM could bind simultaneously to the BMP ligand and the type I and type II receptors, thereby enhancing the formation of the active BMP signalling complex[102]. However, as discussed in Chapter 4, RGMB and BMPR1A appear to compete for ligand binding in solution (Fig. 4.16). In theory, an RGM-BMP2-BMPR1A-BMPR2 complex could form around the sym-

metric BMP2 dimer *in vivo*, particularly given the similar binding affinities of eBMPR1A and eRGMB for BMP2 (Fig. 4.13b and 4.15a). FRET-FLIM studies to show the close association of fluorescently tagged RGM and BMPR1A upon the addition of the BMP ligand would indicate concurrent ligand binding. As RGM is a GPI-anchored protein, attaching a C-terminal intracellular tag might prove problematic; the RGM ectodomain could potentially be fused to a trans-membrane helix, but this may alter the behaviour of the protein within the cell membrane.

In another study RGMA was shown to increase the utilisation of the type II receptor ACVR2A by BMP2/4, which normally only signal via BMPR2[189]. The authors propose a model whereby binding of RGMA to BMP2 favours its association with ACVR2A over BMPR2. As described previously, the ACVR2A ectodomain appears to be accommodated in the RGM_{ND}-BMP2 complex (Fig. 4.16) and could function in the endocytosis-linked model for RGM-activated BMP signalling. It is currently unclear whether an RGM-BMP2-ACVR2A complex can form in solution, although a superposition of the RGM_{CND}-BMP2 complex with the published eBMPR1A-BMP2 (PDB ID: 1ES7[241]) (Fig. 6.3a) and ternary eBMPR1A-BMP2-eACVR2A (PDB ID: 2GOO[298]) (Fig. 6.3b) complexes using the BMP2 ligand as the reference demonstrates that RGM and ACVR2A binding may not be compatible. In the eBMPR1A-BMP2 structure the BMP2 finger loop region adopts a straightened conformation compared to the orientation of this finger in the RGM_{ND}-BMP2 structure where it bends upwards towards the RGM. This difference in conformation is also seen between the RGM_{ND}-BMP2 and the ternary eBMPR1A-BMP2-eACVR2A structure. It may be that the flatter orientation of the BMP2 finger is required for ACVR2A binding, in which case RGM binding would disrupt this; still a MALS experiment could confirm whether this ternary complex is able to form in solution.

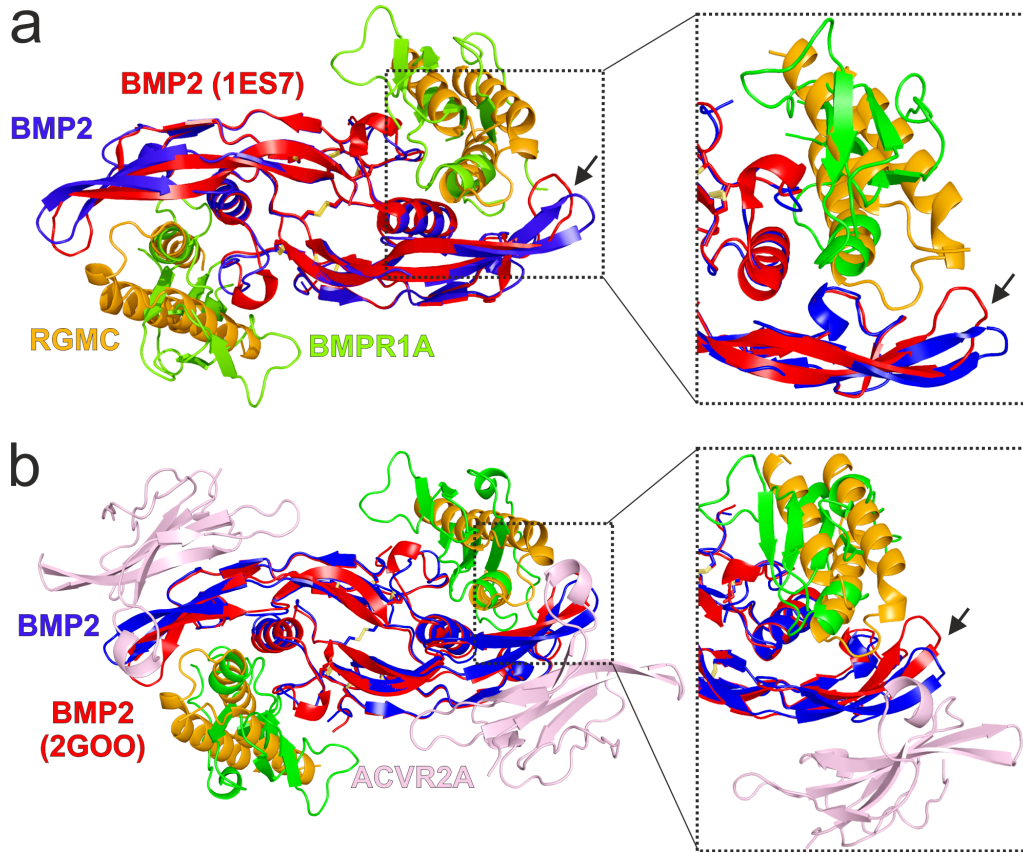


Figure 6.3: Superposition of the RGM_{CND}-BMP2 complex with the eBMPR1A-BMP2 and eBMPR1A-BMP2-eACVR2A complexes. (a-b) Left panels show a cartoon representation of a superposition of the (a) RGM_{CND}-BMP2 and eBMPR1A-BMP2 or (b) RGM_{CND}-BMP2 and eBMPR1A-BMP2-eACVR2A complexes using BMP2 as reference. BMP2 from the RGM_{CND}-BMP2 complex is shown in blue whereas BMP2 from the published complexes is shown in red. RGM_{CND}= orange, eBMPR1A = green and eACVR2A = pink. Right panels show a close-up view of the BMP2 finger region in the complex, arrows highlight the area where the loop conformations differ.

6.4 A model for RGM-mediated NEO1 and BMP signalling

In this thesis, models for RGM signalling through both the NEO1 and BMP pathways have been proposed based on the structural, biophysical and cellular data presented. In Chapter 3 a model for RGM signalling through NEO1 was proposed in which NEO1 pre-clusters on the membrane and then RGM, through

cis (same-cell) or *trans* (cell-to-cell) interactions, acts a molecular staple bringing together and fixing the orientation between two NEO1 receptors to form signalling competent dimers (Fig. 3.18).

In the proposed model of RGM-activated BMP signalling discussed in Chapter 4 (Fig. 4.24), RGM binds to the BMP ligand through a *cis* interaction and this complex is internalised. Upon gradual acidification of milieu the RGM-BMP2 complex dissociates in the endosome allowing the BMP ligand to bind to the internalised pool of type I and type II receptors to activate a higher level of signalling from the endosomal platform. In this model, endocytosis of the BMP ligand through its interaction with RGM and subsequent dissociation of the complex is implicated as key to the activation of BMP signalling. The RGM-BMP2 interaction was shown to be pH-sensitive and to dissociate below pH 6.5, which supports this model (Fig. 4.17). However, despite using a number of different fluorescence microscopy-based approaches (Fig. 4.19-4.23), data supporting the endocytosis of the RGM-BMP complex has yet to be obtained.

RGM is a GPI-anchored protein and as such does not have a cytosolic domain that can be important for endocytic targeting. Nevertheless, this does not necessarily preclude the internalisation of RGMs, for example GPI-anchored proteins have been found in several types of endosomes and appear to be specifically internalised through a dynamin-independent pathway in various mammalian cell-types[333, 334, 335, 336]. It is therefore possible for RGMs to independently mediate the internalisation of the BMP ligand. However, given the data outlined in Chapter 5 (Fig. 5.10 and 5.11), describing the clustering of a BMP2-RGMB-NEO1 complex on the membrane it is an attractive proposition that the NEO1-RGM signalling hub could mediate the internalisation of the BMP ligand. The localisation of both NEO1 and the BMP ligand to membrane micro-domains has been shown to be important for signalling[198, 197]. Interestingly, GPI-anchored proteins have been shown to accumulate in membrane micro-domains,

as the GPI-anchor itself has been shown to confer association with detergent-resistant membrane fractions known as lipid rafts[337, 338, 339, 340, 341]. Even though these structures appear to be very small (sub-micron[342]), clusters of RGMs in these micro-domains could act as nucleation sites for NEO1 receptors and BMP ligands, which would then convert these clusters into ordered arrays consisting of NEO1 and BMP dimers linked by RGMs (Fig. 5.9).

NEO1 and BMP signalling are known to be interdependent and the models proposed in this thesis for RGM-NEO1 and RGM-BMP signalling could be merged to show how, in the presence of all three components, both NEO1 and BMP signalling could be activated (Fig. 6.4). The BMP-induced arrays of signalling competent RGM-NEO1 dimers could amplify the NEO1 signal compared to the signalling level in the absence of BMP ligand. Following this, it is plausible that NEO1, through interactions between its cytosolic domain with components of the clathrin-mediated endocytic pathway, could direct these clusters for internalisation. In the low pH environment of the endosome the pH sensitive RGM-BMP interaction is disrupted, allowing BMP to bind to its receptors and activating the enhanced endosome-specific signal. Concurrently, due to the pH-dependence of the RGM-NEO1 site-2 interaction (Fig. 3.14), the 2:2 RGM-NEO1 complex also dissociates in the endosome, terminating the NEO1 signal. In this model for RGM-mediated BMP and NEO1 signalling the internalisation of the BMP-RGM-NEO1 arrays serves the dual purpose of activating BMP signalling and also an efficient signal termination mechanism for NEO1 signalling. An interesting side-effect of this proposed model would be sequential and binary NEO1 and BMP signalling, both components are required for fully activated signalling but NEO1 signalling would occur first followed by BMP signalling. This model is presently pure conjecture, but an analysis of NEO1 endocytosis in the presence and absence of both BMP2 and RGM perhaps using fluorescence microscopy would begin to determine whether this it is realistic.

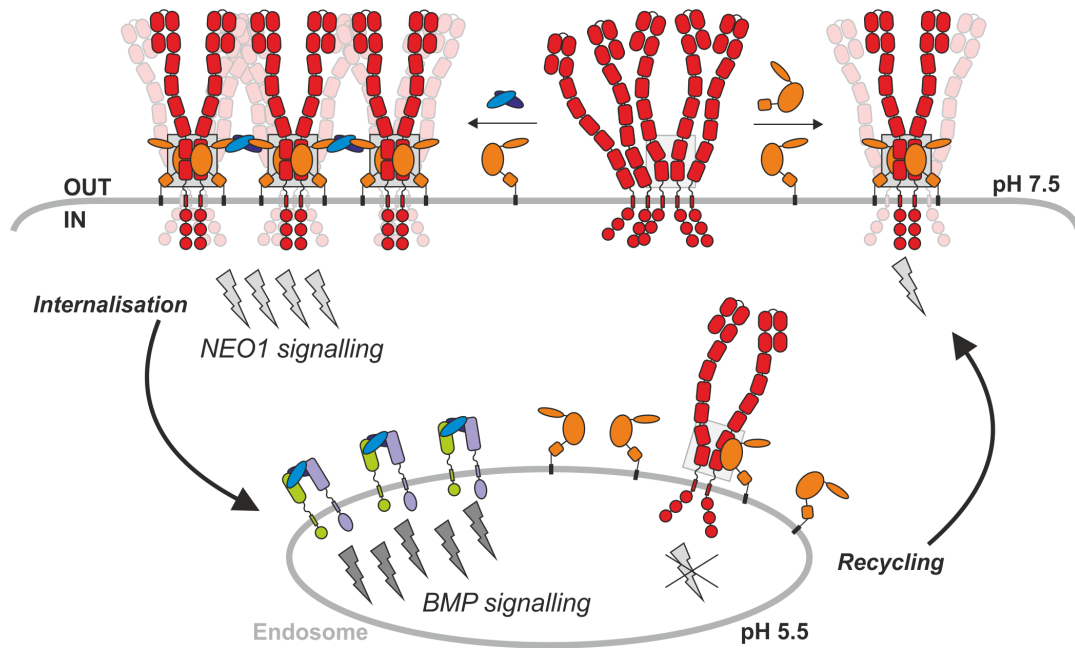


Figure 6.4: A model for RGM-mediated NEO1 and BMP signalling. RGM binding to NEO1 converts pre-clustered receptors into signalling competent NEO1 dimers. In the presence of BMPs these dimers are converted into signalling competent arrays for an enhanced NEO1 signal. The NEO1-RGM signalling hub primes the BMP ligand for internalisation. In the low pH environment of the endosome RGM dissociates from BMP allowing the BMP receptors to bind and activate enhanced BMP signalling. Concurrently the RGM-NEO1 site-2 interaction is disrupted terminating the NEO1 signal. Components are subsequently recycled to the membrane. Membrane = grey line, NEO1 = red, RGM = orange (GPI anchor = black box), BMP2 = blue/cyan, BMPR1A = green, ACVR2A = purple, NEO1 signalling = pale grey lightning bolts, BMP signalling = dark grey lightning bolts, signalling competent NEO1 dimer = shaded grey box.

6.5 Endocytosis as a signalling mechanism in other pathways

The proposed models for RGM, NEO1 and BMP signalling described in this thesis all have endocytosis as a central theme and as a process capable of regulating specific downstream responses of signalling pathways. This is not a novel idea and using the $TGF\beta$ and EGF signalling pathways as examples I aim to illustrate the complex regulative potential of incorporating endocytosis into a signalling

model. This shows how viable the involvement of internalisation is in such fundamental signalling systems as NEO1 and BMP signalling and how RGM with its GPI anchor and its ability to bridge the two systems might facilitate this.

Clathrin-mediated endocytosis is important in the regulation of EGF signalling; in the removal and down-regulation of the active signalling complex but also in controlling the specific downstream signals transmitted by this pathway. EGF ligand binding to EGF receptors (EGFR) triggers receptor dimerisation and phosphorylation at several tyrosine residues in the EGFR cytoplasmic domain[343, 344]. This results in the recruitment of phosphotyrosine-binding downstream components, which initiate several signal transduction cascades including the MAPK, Akt and JNK pathways[343, 344]. Crucially ligand binding also directs the recruitment of EGFR to clathrin-coated pits for internalisation[345]. A K44A dynamin mutant expressing cell line (K44A cells), defective in clathrin-dependent endocytosis was used to show that endocytosis is important for the regulation of EGF signalling[346, 318]. Clathrin-mediated endocytosis of EGFR upon EGF binding was potently inhibited in K44A cells and this resulted in enhanced EGF-dependent proliferation as measured both by the MMT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-tetrazolium bromide] cell proliferation assay or ³H-thymidine DNA incorporation assay[318]. This suggests that clathrin-mediated endocytosis of EGFR is critical for attenuating EGF signalling, which is similar to the model proposed for RGM-activated NEO1 signalling termination. In addition, analysis of the levels of tyrosine phosphorylation in downstream targets of EGF signalling demonstrated that several proteins were hyper- or hypo-phosphorylated in K44A cells[318]. For example, the level of EGFR tyrosine phosphorylation was reduced in K44A cells suggesting that endocytosis is required for full EGFR phosphorylation[347, 318]. Similarly, activation of downstream MAP kinases ERK1 and ERK2 was suppressed in the endocytosis defective cells suggesting that this trafficking is important for activation of the MAPK pathway[318].

Therefore, endocytosis regulates the specific downstream signalling initiated by the EGF pathway; it may be then that RGM-targeted endosomal BMP signalling activates a different subset of downstream signalling cascades.

Endocytosis, and particularly the balance between clathrin-mediated and caveolae-dependent internalisation, is also important for regulation of the $\text{TGF}\beta$ signalling pathway. Similar to the BMP pathway, the $\text{TGF}\beta$ ligand assembles a hexaheteromeric complex comprising the $\text{TGF}\beta$ receptors $\text{TGF}\beta\text{R1}$ and $\text{TGF}\beta\text{R2}$ [300, 293]. $\text{TGF}\beta\text{R2}$ transphosphorylates and activates $\text{TGF}\beta\text{R1}$ [293], which then activates the downstream receptor-activated SMAD2/3[348], a process facilitated by the SMAD-anchor protein SARA (SMAD anchor for receptor activation)[349]. SMAD2/3 then associate with SMAD4 and this complex translocates to the nucleus to affect gene transcription. $\text{TGF}\beta$ receptors were shown to internalise through both clathrin-dependent and caveolae-mediated routes[179]. Three colour immunofluorescence studies showed co-localisation of haemagglutinin (HA)-tagged receptors with both EEA1, a marker of the early endosome, a downstream compartment in clathrin-mediated endocytosis; and caveolin-1, a marker of the caveolae-dependent pathway[179]. Electron microscopy of gold-labelled receptors confirmed this analysis and demonstrated that after 5 minutes of internalisation 28 % of the receptors were in caveolae structures and 20 % were in clathrin-coated structures[179]. Similarly, it was shown that BMP receptors are also internalised through these two distinct pathways, although a thorough analysis of this has not yet been completed[180, 181]. Interestingly, biotinylated $\text{TGF}\beta$ ligand was also shown to be internalised through both endocytic pathways but did not appear to affect receptor trafficking, suggesting that receptor endocytosis is constitutive even in the absence of ligand[179]. Endocytosis of the BMP ligand has also been demonstrated[183] and comparably the internalisation of the BMP receptors was also suggested to occur even in the absence of ligand[180].

$\text{TGF}\beta$ receptors are found in both raft and non-raft domains on the plasma mem-

brane and disrupting their association with rafts directs the receptors into the EEA1-positive compartment in the clathrin-mediated internalisation pathway[179]. This suggests not only that TGF β receptors are organised into distinct membrane micro-domains, but also that this localisation regulates the pathway through which they are internalised. In relation to the model proposed for RGM-activated NEO1 and BMP signalling (Fig. 6.4), RGM-mediated targeting of NEO1 and BMP to specific membrane micro-domains may regulate the route of internalisation. An example where a GPI-anchored protein is important in partitioning receptors into specific membrane domains is in GDNF and neuronal cell adhesion molecule p140^{NCAM} signalling; both of these molecules are localised to lipid rafts by the GPI-anchored protein GFR α 1 (GDNF-family-receptor- α 1), which results in activation of signalling[350, 351].

Clathrin- and caveolae-dependent pathways fulfil opposite regulatory functions in TGF β signalling. Trafficking of the raft-associated TGF β receptors to the caveolae-dependent pathway brings them into association with SMAD7-Smurf2 ubiquitin ligase complexes, which targets the receptors for degradation[179]. In contrast, SARA co-localises extensively with EEA1- and Rab5-positive early endosomes in the clathrin-mediated endocytosis pathway, and trafficking of receptors through this endocytic route sequesters receptors away from the caveolae-mediated pathway thereby activating TGF β signalling[179]. Interestingly, like SARA, an endosome-localised SMAD1 anchor protein Endofin was recently shown to enhance BMP signalling[184] and in zebrafish the endosome adaptor protein Eps15R was found to be important for activation of the BMP pathway[185]; this potentially suggests that clathrin-mediated internalisation of the BMP receptors is important for signalling, similar to TGF β .

It is clear then that the different routes of TGF β endocytosis regulate downstream signalling, with caveolae negatively regulating the pathway by targeting the receptors for degradation and the clathrin-mediated route of endocytosis ac-

tivating TGF β signalling by preventing receptor degradation and facilitating the association with SMAD2/3 via SARA. It is interesting to speculate on the potential similarities with an endocytosis-linked mechanism for activation of BMP signalling. BMP receptors are constitutively targeted to the caveolae and clathrin-dependent internalisation pathways in a ligand-independent manner[180, 181]. The balance between these two endocytic routes may be determined by their association with membrane micro-domains, similar to TGF β signalling, which could potentially be regulated by NEO1 and RGM[198, 197]. This may result in an internalised pool of BMP receptors with a bias towards one type of intracellular compartment. The activation of signalling is mediated through the BMP ligand and its level of internalisation, which may be controlled by the RGM-NEO1 interaction. RGM targets the BMP ligand for clathrin-mediated endocytosis delivering it to the pool of internalised receptors in the early endosome, which are protected from degradation and associated with SMAD1/5/8 via anchors such as Endofin. It will be interesting to investigate whether the caveolae-dependent route of endocytosis results in degradation of BMP receptors and whether RGMs enhance the delivery of BMP to endosomal compartments via clathrin-mediated endocytosis.

Overall, it is clear that dynamic regulation of endocytosis is important in many signalling pathways, both as a mechanism of signal termination and also to control the nature of the molecular response. The early endosome may potentially represent a signalling centre for both the TGF β and BMP pathways that functions to sequester active receptor complexes and act as a nucleation site for downstream signalling components such as SMADs.

6.6 RGMs as targets for clinical interventions

Aberrant RGM function has been linked to a number of diverse disease states including cancer[126, 128, 127], MS[79], and iron homeostasis disorders[107, 102, 54]. This diversity is linked to the wide range of functions mediated by the three RGM family members, which is a direct result of their ability to signal through both the BMP and NEO1 signalling pathways. Blocking the action of RGMs has been shown to have clinical potential in many different systems, for example, a knockdown of RGMB in a mouse model of colorectal cancer significantly slowed down tumour growth[128]. Similarly, antibodies against RGMA reduced the immune response both in a mouse model of MS and in cell samples from MS patients[79, 82]. RGMA antibodies also showed potential in the treatment of paralysis following CNS damage; in rats these antibodies significantly improved motility following spinal cord injury[67]. Furthermore, both RGMC-specific antibodies and soluble Fc-RGMC have been shown to increase serum iron levels in mice and monkeys, highlighting their potential in treating anaemia of inflammation in humans[103, 114]. Another recently identified role of RGMB is in the development of respiratory tolerance in the lung, a state of immunological non-responsiveness induced by exposure to innocuous inhaled antigens, which is mediated through the RGMB-PD-L2 interaction[332]. Interestingly, antibodies against RGMB inhibited the development of respiratory tolerance and led to increased T-cell proliferation and IL-4 expression compared to control antibodies in mice[332]. The authors of this study suggest that these signalling effects triggered by RGMB and PD-L2 are most likely mediated through RGMB binding partners NEO1 and/or BMP. Targeting the PD-L2 pathway, potentially through RGMs, has therapeutic potential in the treatment of asthma. Furthermore, antibodies against another PD-L2 receptor PD-1 are a promising cancer immunotherapy and have been tested in clinical trials; however, lung inflammation leading to pneumonitis was a severe side effect potentially caused by the

action of PD-L2[352, 332]. Targeting the RGMB-PD-L2 interaction may enhance PD-1 targeted cancer immunotherapy by minimising adverse effects.

Prior to the work outlined in this thesis the structure of RGMs was unknown, now the high-resolution crystal structures of the RGM_{ND}-BMP2 (Fig. 6.5a), eRGMB-NEO1_{FN56} (Fig. 6.5d) and BMP2-eRGMB-NEO1_{FN56} complexes have revealed the modular domain architecture of the RGM family. These structures have also determined the location of the RGM-BMP and RGM-NEO1 binding sites and provided atomic resolution details of the interfaces. So far inhibiting the action of RGMs has been achieved using antibodies, however, with the availability of this structural data it could be possible to perform structure-guided drug design to develop novel therapeutics targeting RGMs. An analysis of the RGM-BMP and RGM-NEO1 binding interfaces show several cavities and pockets that could potentially be targeted in fragment-based screening and small molecule docking experiments (Fig. 6.5b,e). Small-molecule drugs have the advantage of being easier to administer and cheaper to produce than antibody therapies and often have increased levels of tissue and organ penetration making them potentially clinically useful entities.

Another strategy that has already been implemented to inhibit BMP signalling is the use of “ligand traps”. These are proteins typically containing the ligand-binding domain of a receptor coupled to an immunoglobulin Fc domain. These proteins then have high binding affinity for the ligand coupled with increased stability from the Fc tag for use *in vivo*. They bind to the ligand of interest and occlude binding to the endogenous receptors thereby inhibiting signalling. Ligand traps including ALK1-Fc and ACVR2A-Fc targeting BMP signalling have been tested as cancer therapeutics in clinical trials[353, 354] and BMPR1A-Fc[355] and ACVR2B-Fc[356] have both shown promise as BMP signalling antagonists but have yet to proceed into the clinic. A comparison of the RGM_{CND}-BMP2 and eBMPR1A-BMP2 complex structures showed that RGM binds to the BMP

ligand at the same site at the type I receptor and an SPR experiment showed that the two proteins compete for ligand-binding, suggesting that RGM_{ND} also has the potential to act as a ligand-trap to inhibit BMP signalling. However, MALS experiments showed that the RGMB-BMP interaction is pH dependent which was surmised to be the reason that soluble eRGMB did not inhibit BMP signalling in the same manner as eBMPRI1A. Therefore, in order to make RGM_{ND} a functional ligand-trap the source of the pH dependence would need to be established and mutations made to abolish this characteristic. Despite this requirement for further research, the RGM N-terminal domain takes the form of a triple helix bundle stabilised by three disulphide bonds (Fig. 6.5c), which is a highly stable, compact module that would be ideally suited to being adapted as a ligand trap, and the expression and purification strategies outlined in this thesis provide a robust protocol for obtaining milligram quantities with >95 % purity of this BMP-binding domain. RGM_{CND} would be potentially useful in treating anaemia of inflammation as well as for cancer therapies.

This thesis contains structural and sequence-based analysis that, together with the eRGMB-NEO1_{FN56} structure, comprises the first definition of the RGM domain boundaries. This analysis facilitated cloning of constructs to express the separate domains and for the first time RGM proteins could be produced in a modular fashion in milligram quantities. Stable constructs of both RGM_{ND} and RGM_{ΔND} could be expressed and purified (Fig. 6.5c,f) and these could be useful in the generation of antibodies or nanobodies targeting either the RGM N- or C-terminal domain, for example to specifically disrupt the RGM-NEO1 or the RGM-BMP interaction (Fig. 6.5a,c). These antibodies could have the advantage of fewer off-target effects by only inhibiting one part of the RGM signalling system; although more research is required to determine to what extent NEO1 and BMP signalling are interdependent before it is certain that one pathway can be specifically targeted at all.

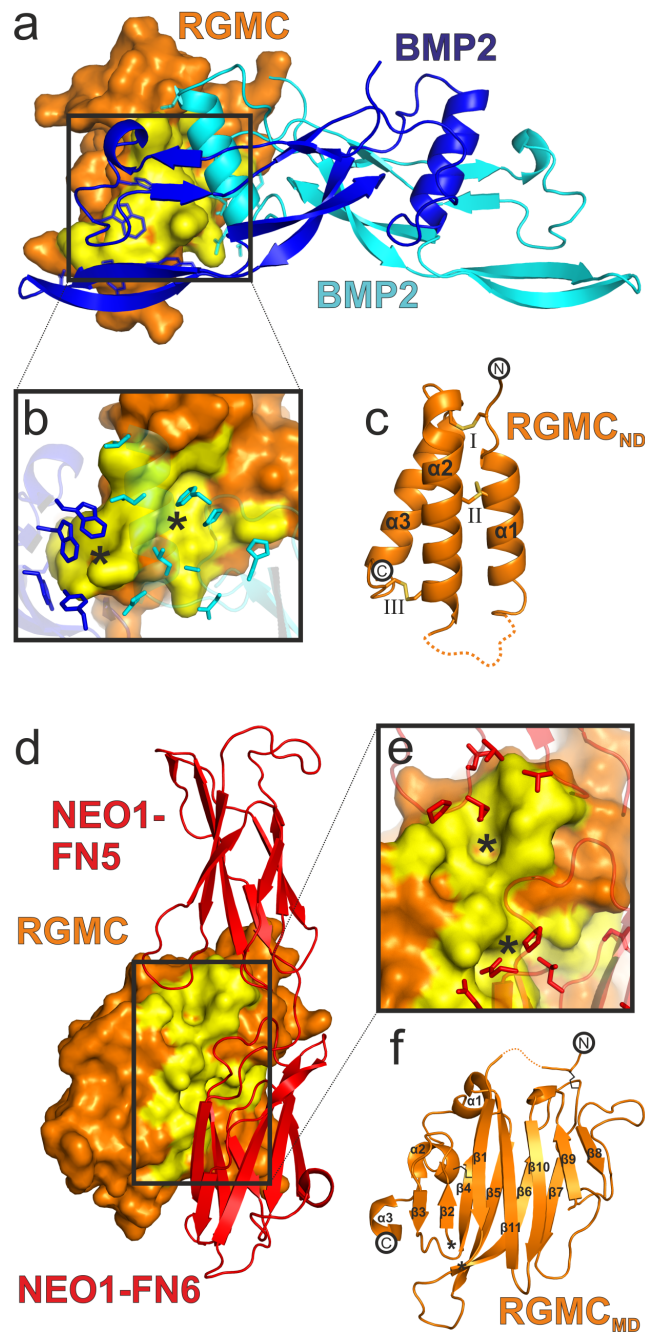


Figure 6.5: RGM_{ND}-BMP2 and eRGM-NEO1_{FN56} complex structures could aid the design of targeted therapeutics. (a) The RGM_{ND}-BMP2 complex. RGM surface shown in orange, interacting region shown in yellow, BMP2 (blue, cyan) shown in ribbon representation. (b) Close-up of the RGM-BMP2 interface, binding cavities in the RGM surface highlighted with asterisks (*). BMP2 residues shown as sticks. (c) RGM_{ND} ribbon representation, disulphide bonds shown as sticks. (d) The eRGM-NEO1_{FN56} complex. RGM representation as in (a), NEO1 (red) shown in ribbon representation. (e) Close-up of the eRGM-NEO1_{FN56} interface, binding cavities in the RGM surface highlighted with asterisks (*). NEO1 residues shown as sticks. (f) RGM middle domain ribbon representation, disulphide bonds shown as sticks, auto-catalytic cleavage site highlighted by asterisks (*).

6.7 Conclusions

In this thesis I have described models for RGM signalling through NEO1 and RGM-activated BMP signalling, both of which have been informed by structural, biophysical and cellular data. In Chapter 3, I described how RGM acts as a molecular staple bringing together and fixing the orientation of two NEO1 receptors to trigger downstream signalling. Furthermore, the pH dependence of the RGM-NEO1 2:2 complex suggested an endocytosis-linked mechanism of signal attenuation, which is mimicked in other systems such as the EGF pathway. In Chapter 4, I outlined how RGM and BMPRII likely compete for BMP ligand binding *in vivo* and, together with the pH dependence of the RGM-BMP interaction, this suggested an endocytosis-linked mechanism for RGM-activated BMP signalling dependent on the subcellular localisation of the BMP receptors. In addition, in Chapter 5, I presented data that demonstrated for the first time that RGM forms a physical bridge linking NEO1 and BMP, which results in RGM-mediated clustering of NEO1 and BMP dimers that could potentially be important for signalling.

The work presented in this thesis has also revealed the molecular rationale behind two groups of JHH-linked RGMC mutations. Mutations in the middle domain of RGMC abolish the secretion of the protein whereas mutations in the N-terminal domain of RGMC disrupt binding to the BMP ligand. Ultimately both types of mutations prevent RGMC-activation of BMP6 signalling, resulting in blood iron overload. This thesis describes the first structural insights into the RGM family and provides atomic resolution details of the interactions of RGMs with NEO1 and BMPs. As outlined above, these novel structural insights may facilitate structure-guided drug design to target RGM binding interfaces for the treatment of RGM-linked diseases and conditions for which RGM antibodies are already being trialled, including MS, anaemia of inflammation and regeneration

of neuronal growth.

These data together have opened up many exciting new avenues for research mostly regarding the subcellular localisation and trafficking of the signalling components; in particular, linking RGMs to BMP endocytosis and characterising this in different biological contexts. The translocation of the entire signalling machinery to place it in close proximity to the nucleus is a very efficient way of targeting changes in gene expression. Further work is required to establish the endosome as an essential signalling platform for BMPs, as it is for TGF β , EGF and GDNF signalling, however the proposed endocytosis-linked mechanism for RGM-activated BMP signalling, involving the internalisation of the active signalling complex may represent a paradigm for many “cell-surface” signalling events. The importance of the RGM signalling hub is just beginning to emerge, and given their wealth of physiological roles there is no doubt that the repertoire of RGM binding partners and signalling modes will continue to expand.

Bibliography

- [1] Wolpert, L. Positional information and the spatial pattern of cellular differentiation. *J Theor Biol* **25**, 1–47 (1969).
- [2] Turing, A. M. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. **237**, 37–72 (1952).
- [3] Nusslein-Volhard, C., Lohs-Schardin, M., Sander, K. & Cremer, C. A dorso-ventral shift of embryonic primordia in a new maternal-effect mutant of drosophila. *Nature* **283**, 474–6 (1980).
- [4] Nusslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in drosophila. *Nature* **287**, 795–801 (1980).
- [5] Massague, J. & Chen, Y. G. Controlling tgf-beta signaling. *Genes Dev* **14**, 627–44 (2000).
- [6] Szebenyi, G. & Fallon, J. F. Fibroblast growth factors as multifunctional signaling factors. *Int Rev Cytol* **185**, 45–106 (1999).
- [7] Betsholtz, C., Karlsson, L. & Lindahl, P. Developmental roles of platelet-derived growth factors. *Bioessays* **23**, 494–507 (2001).
- [8] Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev* **15**, 3059–87 (2001).
- [9] Nusse, R. & Varmus, H. E. Wnt genes. *Cell* **69**, 1073–87 (1992).
- [10] Gilbert, S. F. *Developmental Biology* (Sinauer, Sunderland (MA), 2013), tenth edn.
- [11] Sadava, D. E., Hillis, D. M., Heller, H. C. & Berenbaum, M. *LIFE: The Science of Biology* (Sinauer Associates, New York, 2013), tenth edn.
- [12] Spemann, H. & Mongold, H. Über induktion von embryonalagen durch implantation artfremder organisatoren. *Roux' Arch. Entw. Mech.* **100**, 599–638 (1924).
- [13] De Robertis, E. M. Spemann's organizer and self-regulation in amphibian embryos. *Nat Rev Mol Cell Biol* **7**, 296–302 (2006).

- [14] Streit, A., Berliner, A. J., Papanayotou, C., Sirulnik, A. & Stern, C. D. Initiation of neural induction by fgf signalling before gastrulation. *Nature* **406**, 74–8 (2000).
- [15] Wilson, S. I., Graziano, E., Harland, R., Jessell, T. M. & Edlund, T. An early requirement for fgf signalling in the acquisition of neural cell fate in the chick embryo. *Curr Biol* **10**, 421–9 (2000).
- [16] Baker, J. C., Beddington, R. S. & Harland, R. M. Wnt signaling in xenopus embryos inhibits bmp4 expression and activates neural development. *Genes Dev* **13**, 3149–59 (1999).
- [17] Greene, N. D. & Copp, A. J. Neural tube defects. *Annu Rev Neurosci* **37**, 221–42 (2014).
- [18] Chilton, J. K. Molecular mechanisms of axon guidance. *Dev Biol* **292**, 13–24 (2006).
- [19] Kalil, K. & Dent, E. W. Touch and go: guidance cues signal to the growth cone cytoskeleton. *Curr Opin Neurobiol* **15**, 521–6 (2005).
- [20] Gupton, S. L. & Gertler, F. B. Filopodia: the fingers that do the walking. *Sci STKE* **2007**, re5 (2007).
- [21] Tanaka, E. & Kirschner, M. W. The role of microtubules in growth cone turning at substrate boundaries. *J Cell Biol* **128**, 127–37 (1995).
- [22] Tanaka, E. M. & Kirschner, M. W. Microtubule behavior in the growth cones of living neurons during axon elongation. *J Cell Biol* **115**, 345–63 (1991).
- [23] Burnette, D. T. *et al.* Myosin ii activity facilitates microtubule bundling in the neuronal growth cone neck. *Dev Cell* **15**, 163–9 (2008).
- [24] Geraldo, S., Khanzada, U. K., Parsons, M., Chilton, J. K. & Gordon-Weeks, P. R. Targeting of the f-actin-binding protein drebrin by the microtubule plus-tip protein eb3 is required for neuritogenesis. *Nat Cell Biol* **10**, 1181–9 (2008).
- [25] Medeiros, N. A., Burnette, D. T. & Forscher, P. Myosin ii functions in actin-bundle turnover in neuronal growth cones. *Nat Cell Biol* **8**, 215–26 (2006).
- [26] Gallo, G. Rhoa-kinase coordinates f-actin organization and myosin ii activity during semaphorin-3a-induced axon retraction. *J Cell Sci* **119**, 3413–23 (2006).
- [27] Gallo, G., Yee, J., H. F. & Letourneau, P. C. Actin turnover is required to prevent axon retraction driven by endogenous actomyosin contractility. *J Cell Biol* **158**, 1219–28 (2002).

- [28] Wylie, S. R. & Chantler, P. D. Myosin iia drives neurite retraction. *Mol Biol Cell* **14**, 4654–66 (2003).
- [29] Kozma, R., Ahmed, S., Best, A. & Lim, L. The ras-related protein cdc42hs and bradykinin promote formation of peripheral actin microspikes and filopodia in swiss 3t3 fibroblasts. *Mol Cell Biol* **15**, 1942–52 (1995).
- [30] Nobes, C. D. & Hall, A. Rho, rac, and cdc42 gtpases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. *Cell* **81**, 53–62 (1995).
- [31] Ridley, A. J. & Hall, A. The small gtp-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* **70**, 389–99 (1992).
- [32] Ridley, A. J., Paterson, H. F., Johnston, C. L., Diekmann, D. & Hall, A. The small gtp-binding protein rac regulates growth factor-induced membrane ruffling. *Cell* **70**, 401–10 (1992).
- [33] Luo, Y., Raible, D. & Raper, J. A. Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones. *Cell* **75**, 217–27 (1993).
- [34] Drescher, U. *et al.* In vitro guidance of retinal ganglion cell axons by rags, a 25 kda tectal protein related to ligands for eph receptor tyrosine kinases. *Cell* **82**, 359–70 (1995).
- [35] Placzek, M., Tessier-Lavigne, M., Jessell, T. & Dodd, J. Orientation of commissural axons in vitro in response to a floor plate-derived chemoattractant. *Development* **110**, 19–30 (1990).
- [36] Monnier, P. P. *et al.* Rgm is a repulsive guidance molecule for retinal axons. *Nature* **419**, 392–5 (2002).
- [37] Wilson, N. H. & Key, B. Neogenin interacts with rgma and netrin-1 to guide axons within the embryonic vertebrate forebrain. *Dev Biol* **296**, 485–98 (2006).
- [38] Hall, A. C., Lucas, F. R. & Salinas, P. C. Axonal remodeling and synaptic differentiation in the cerebellum is regulated by wnt-7a signaling. *Cell* **100**, 525–35 (2000).
- [39] McMahon, J. A. *et al.* Noggin-mediated antagonism of bmp signaling is required for growth and patterning of the neural tube and somite. *Genes Dev* **12**, 1438–52 (1998).
- [40] Friedl, P. & Wolf, K. Plasticity of cell migration: a multiscale tuning model. *J Cell Biol* **188**, 11–9 (2010).
- [41] Geraldo, S. & Gordon-Weeks, P. R. Cytoskeletal dynamics in growth-cone steering. *J Cell Sci* **122**, 3595–604 (2009).

- [42] Mueller, B. K., Yamashita, T., Schaffar, G. & Mueller, R. The role of repulsive guidance molecules in the embryonic and adult vertebrate central nervous system. *Philos Trans R Soc Lond B Biol Sci* **361**, 1513–29 (2006).
- [43] Cox, E. C., Muller, B. & Bonhoeffer, F. Axonal guidance in the chick visual system: posterior tectal membranes induce collapse of growth cones from the temporal retina. *Neuron* **4**, 31–7 (1990).
- [44] Walter, J., Kern-Veits, B., Huf, J., Stolze, B. & Bonhoeffer, F. Recognition of position-specific properties of tectal cell membranes by retinal axons in vitro. *Development* **101**, 685–96 (1987).
- [45] Walter, J., Henke-Fahle, S. & Bonhoeffer, F. Avoidance of posterior tectal membranes by temporal retinal axons. *Development* **101**, 909–13 (1987).
- [46] Stahl, B., Muller, B., von Boxberg, Y., Cox, E. C. & Bonhoeffer, F. Biochemical characterization of a putative axonal guidance molecule of the chick visual system. *Neuron* **5**, 735–43 (1990).
- [47] Walter, J., Muller, B. & Bonhoeffer, F. Axonal guidance by an avoidance mechanism. *J Physiol (Paris)* **84**, 104–10 (1990).
- [48] Muller, B. K., Jay, D. G. & Bonhoeffer, F. Chromophore-assisted laser inactivation of a repulsive axonal guidance molecule. *Curr Biol* **6**, 1497–502 (1996).
- [49] Ruoslahti, E. Rgd and other recognition sequences for integrins. *Annu Rev Cell Dev Biol* **12**, 697–715 (1996).
- [50] Sadler, J. E. Biochemistry and genetics of von willebrand factor. *Annu Rev Biochem* **67**, 395–424 (1998).
- [51] Niederkofer, V., Salie, R., Sigrist, M. & Arber, S. Repulsive guidance molecule (rgm) gene function is required for neural tube closure but not retinal topography in the mouse visual system. *J Neurosci* **24**, 808–18 (2004).
- [52] Schmidtmer, J. & Engelkamp, D. Isolation and expression pattern of three mouse homologues of chick rgm. *Gene Expr Patterns* **4**, 105–10 (2004).
- [53] Samad, T. A. *et al.* Dragon: a member of the repulsive guidance molecule-related family of neuronal- and muscle-expressed membrane proteins is regulated by drg11 and has neuronal adhesive properties. *J Neurosci* **24**, 2027–36 (2004).
- [54] Papanikolaou, G. *et al.* Mutations in hfe2 cause iron overload in chromosome 1q-linked juvenile hemochromatosis. *Nat Genet* **36**, 77–82 (2004).
- [55] Oldekamp, J., Kramer, N., Alvarez-Bolado, G. & Skutella, T. Expression pattern of the repulsive guidance molecules rgm a, b and c during mouse development. *Gene Expr Patterns* **4**, 283–8 (2004).

- [56] Metzger, M., Conrad, S., Alvarez-Bolado, G., Skutella, T. & Just, L. Gene expression of the repulsive guidance molecules during development of the mouse intestine. *Dev Dyn* **234**, 169–75 (2005).
- [57] Xia, Y. *et al.* Localization and action of dragon (repulsive guidance molecule b), a novel bone morphogenetic protein coreceptor, throughout the reproductive axis. *Endocrinology* **146**, 3614–21 (2005).
- [58] Lidell, M. E., Johansson, M. E. & Hansson, G. C. An autocatalytic cleavage in the c terminus of the human muc2 mucin occurs at the low ph of the late secretory pathway. *J Biol Chem* **278**, 13944–51 (2003).
- [59] Xiong, J. P. *et al.* Crystal structure of the extracellular segment of integrin alpha vbeta3 in complex with an arg-gly-asp ligand. *Science* **296**, 151–5 (2002).
- [60] Matsunaga, E., Nakamura, H. & Chedotal, A. Repulsive guidance molecule plays multiple roles in neuronal differentiation and axon guidance. *J Neurosci* **26**, 6082–8 (2006).
- [61] Gessert, S., Maurus, D. & Kuhl, M. Repulsive guidance molecule a (rgm a) and its receptor neogenin during neural and neural crest cell development of xenopus laevis. *Biol Cell* **100**, 659–73 (2008).
- [62] Brinks, H. *et al.* The repulsive guidance molecule rgma is involved in the formation of afferent connections in the dentate gyrus. *J Neurosci* **24**, 3862–9 (2004).
- [63] Lah, G. J. & Key, B. Dual roles of the chemorepellent axon guidance molecule rgma in establishing pioneering axon tracts and neural fate decisions in embryonic vertebrate forebrain. *Dev Neurobiol* **72**, 1458–70 (2012).
- [64] Metzger, M., Conrad, S., Skutella, T. & Just, L. Rgma inhibits neurite outgrowth of neuronal progenitors from murine enteric nervous system via the neogenin receptor in vitro. *J Neurochem* **103**, 2665–78 (2007).
- [65] O’Leary, C. *et al.* Rgma regulates cortical interneuron migration and differentiation. *PLoS One* **8**, e81711 (2013).
- [66] Tassew, N. G., Charish, J., Seidah, N. G. & Monnier, P. P. Ski-1 and furin generate multiple rgma fragments that regulate axonal growth. *Dev Cell* **22**, 391–402 (2012).
- [67] Hata, K. *et al.* Rgma inhibition promotes axonal growth and recovery after spinal cord injury. *J Cell Biol* **173**, 47–58 (2006).
- [68] Schwab, J. M. *et al.* Spinal cord injury-induced lesional expression of the repulsive guidance molecule (rgm). *Eur J Neurosci* **21**, 1569–76 (2005).
- [69] Schwab, J. M. *et al.* Central nervous system injury-induced repulsive guidance molecule expression in the adult human brain. *Arch Neurol* **62**, 1561–8 (2005).

- [70] Brugeaud, A., Tong, M., Luo, L. & Edge, A. S. Inhibition of repulsive guidance molecule, rgma, increases afferent synapse formation with auditory hair cells. *Dev Neurobiol* **74**, 457–66 (2014).
- [71] Rao, Y., Wong, K., Ward, M., Jurgensen, C. & Wu, J. Y. Neuronal migration and molecular conservation with leukocyte chemotaxis. *Genes Dev* **16**, 2973–84 (2002).
- [72] Baggiolini, M. Chemokines and leukocyte traffic. *Nature* **392**, 565–8 (1998).
- [73] Luster, A. D. Chemokines–chemotactic cytokines that mediate inflammation. *N Engl J Med* **338**, 436–45 (1998).
- [74] Ly, N. P. *et al.* Netrin-1 inhibits leukocyte migration in vitro and in vivo. *Proc Natl Acad Sci U S A* **102**, 14729–34 (2005).
- [75] Springer, T. A. Traffic signals for lymphocyte recirculation and leukocyte emigration: the multistep paradigm. *Cell* **76**, 301–14 (1994).
- [76] Suzuki, K. *et al.* Semaphorin 7a initiates t-cell-mediated inflammatory responses through alpha1beta1 integrin. *Nature* **446**, 680–4 (2007).
- [77] Wu, J. Y. *et al.* The neuronal repellent slit inhibits leukocyte chemotaxis induced by chemotactic factors. *Nature* **410**, 948–52 (2001).
- [78] Mirakaj, V. *et al.* Repulsive guidance molecule-a (rgm-a) inhibits leukocyte migration and mitigates inflammation. *Proc Natl Acad Sci U S A* **108**, 6555–60 (2011).
- [79] Muramatsu, R. *et al.* Rgma modulates t cell responses and is involved in autoimmune encephalomyelitis. *Nat Med* **17**, 488–94 (2011).
- [80] Koeberle, P. D., Tura, A., Tassew, N. G., Schlichter, L. C. & Monnier, P. P. The repulsive guidance molecule, rgma, promotes retinal ganglion cell survival in vitro and in vivo. *Neuroscience* **169**, 495–504 (2010).
- [81] Shin, G. J. & Wilson, N. H. Overexpression of repulsive guidance molecule (rgm) a induces cell death through neogenin in early vertebrate development. *J Mol Histol* **39**, 105–13 (2008).
- [82] Tanabe, S. & Yamashita, T. Repulsive guidance molecule-a is involved in th17-cell-induced neurodegeneration in autoimmune encephalomyelitis. *Cell Rep* **9**, 1459–70 (2014).
- [83] Babitt, J. L. *et al.* Repulsive guidance molecule (rgma), a dragon homologue, is a bone morphogenetic protein co-receptor. *J Biol Chem* **280**, 29820–7 (2005).
- [84] Ma, C. H. *et al.* The bmp coreceptor rgmb promotes while the endogenous bmp antagonist noggin reduces neurite outgrowth and peripheral nerve regeneration by modulating bmp signaling. *J Neurosci* **31**, 18391–400 (2011).

- [85] Liu, X. *et al.* Repulsive guidance molecule b inhibits neurite growth and is increased after spinal cord injury. *Biochem Biophys Res Commun* **382**, 795–800 (2009).
- [86] Conrad, S. *et al.* Rgmb controls aggregation and migration of neogenin-positive cells in vitro and in vivo. *Mol Cell Neurosci* **43**, 222–31 (2010).
- [87] Liu, W. *et al.* Dragon (repulsive guidance molecule rgmb) inhibits e-cadherin expression and induces apoptosis in renal tubular epithelial cells. *J Biol Chem* **288**, 31528–39 (2013).
- [88] Xia, Y. *et al.* Dragon (repulsive guidance molecule b) inhibits il-6 expression in macrophages. *J Immunol* **186**, 1369–76 (2011).
- [89] Samad, T. A. *et al.* Dragon, a bone morphogenetic protein co-receptor. *J Biol Chem* **280**, 14122–9 (2005).
- [90] Peiris, D., Pacheco, I., Spencer, C. & MacLeod, R. J. The extracellular calcium-sensing receptor reciprocally regulates the secretion of bmp-2 and the bmp antagonist noggin in colonic myofibroblasts. *Am J Physiol Gastrointest Liver Physiol* **292**, G753–66 (2007).
- [91] Schneeberger, E. E. & Lynch, R. D. The tight junction: a multifunctional complex. *Am J Physiol Cell Physiol* **286**, C1213–28 (2004).
- [92] Xia, Y. *et al.* Dragon enhances bmp signaling and increases transepithelial resistance in kidney epithelial cells. *J Am Soc Nephrol* **21**, 666–77 (2010).
- [93] De Domenico, I., McVey Ward, D. & Kaplan, J. Regulation of iron acquisition and storage: consequences for iron-linked disorders. *Nat Rev Mol Cell Biol* **9**, 72–81 (2008).
- [94] Abboud, S. & Haile, D. J. A novel mammalian iron-regulated protein involved in intracellular iron metabolism. *J Biol Chem* **275**, 19906–12 (2000).
- [95] Donovan, A. *et al.* Positional cloning of zebrafish ferroportin1 identifies a conserved vertebrate iron exporter. *Nature* **403**, 776–81 (2000).
- [96] McKie, A. T. *et al.* A novel duodenal iron-regulated transporter, ireg1, implicated in the basolateral transfer of iron to the circulation. *Mol Cell* **5**, 299–309 (2000).
- [97] Nemeth, E. *et al.* Heparidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. *Science* **306**, 2090–3 (2004).
- [98] Nicolas, G. *et al.* Lack of hepcidin gene expression and severe tissue iron overload in upstream stimulatory factor 2 (usf2) knockout mice. *Proc Natl Acad Sci U S A* **98**, 8780–5 (2001).
- [99] Roetto, A. *et al.* Mutant antimicrobial peptide hepcidin is associated with severe juvenile hemochromatosis. *Nat Genet* **33**, 21–2 (2003).

- [100] Nicolas, G. *et al.* Severe iron deficiency anemia in transgenic mice expressing liver hepcidin. *Proc Natl Acad Sci U S A* **99**, 4596–601 (2002).
- [101] Weinstein, D. A. *et al.* Inappropriate expression of hepcidin is associated with iron refractory anemia: implications for the anemia of chronic disease. *Blood* **100**, 3776–81 (2002).
- [102] Babitt, J. L. *et al.* Bone morphogenetic protein signaling by hemojuvelin regulates hepcidin expression. *Nat Genet* **38**, 531–9 (2006).
- [103] Babitt, J. L. *et al.* Modulation of bone morphogenetic protein signaling in vivo regulates systemic iron balance. *J Clin Invest* **117**, 1933–9 (2007).
- [104] Wang, R. H. *et al.* A role of smad4 in iron metabolism through the positive regulation of hepcidin expression. *Cell Metab* **2**, 399–409 (2005).
- [105] Nemeth, E. *et al.* Il-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin. *J Clin Invest* **113**, 1271–6 (2004).
- [106] Huang, F. W., Pinkus, J. L., Pinkus, G. S., Fleming, M. D. & Andrews, N. C. A mouse model of juvenile hemochromatosis. *J Clin Invest* **115**, 2187–91 (2005).
- [107] Andriopoulos, J., B. *et al.* Bmp6 is a key endogenous regulator of hepcidin expression and iron metabolism. *Nat Genet* **41**, 482–7 (2009).
- [108] Meynard, D. *et al.* Lack of the bone morphogenetic protein bmp6 induces massive iron overload. *Nat Genet* **41**, 478–81 (2009).
- [109] Zhang, A. S., Gao, J., Koeberl, D. D. & Enns, C. A. The role of hepatocyte hemojuvelin in the regulation of bone morphogenic protein-6 and hepcidin expression in vivo. *J Biol Chem* **285**, 16416–23 (2010).
- [110] Lin, L., Goldberg, Y. P. & Ganz, T. Competitive regulation of hepcidin mrna by soluble and cell-associated hemojuvelin. *Blood* **106**, 2884–9 (2005).
- [111] Nili, M., Shinde, U. & Rotwein, P. Soluble repulsive guidance molecule c/hemojuvelin is a broad spectrum bone morphogenetic protein (bmp) antagonist and inhibits both bmp2- and bmp6-mediated signaling and gene expression. *J Biol Chem* **285**, 24783–92 (2010).
- [112] Brasse-Lagnel, C. *et al.* Immunoassay for human serum hemojuvelin. *Haematologica* **95**, 2031–7 (2010).
- [113] Chen, W. *et al.* A novel validated enzyme-linked immunosorbent assay to quantify soluble hemojuvelin in mouse serum. *Haematologica* **98**, 296–304 (2013).

- [114] Boser, P. *et al.* Anti-repulsive guidance molecule c (rgmc) antibodies increases serum iron in rats and cynomolgus monkeys by hepcidin downregulation. *AAPS J* **17**, 930–8 (2015).
- [115] Wakefield, L. M. & Hill, C. S. Beyond tgfbeta: roles of other tgfbeta superfamily members in cancer. *Nat Rev Cancer* **13**, 328–41 (2013).
- [116] Arakawa, H. Netrin-1 and its receptors in tumorigenesis. *Nat Rev Cancer* **4**, 978–87 (2004).
- [117] Fearon, E. R. *et al.* Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* **247**, 49–56 (1990).
- [118] Mazelin, L. *et al.* Netrin-1 controls colorectal tumorigenesis by regulating apoptosis. *Nature* **431**, 80–4 (2004).
- [119] Shin, S. K. *et al.* Epigenetic and genetic alterations in netrin-1 receptors unc5c and dcc in human colon cancer. *Gastroenterology* **133**, 1849–57 (2007).
- [120] Batlle, E. *et al.* Ephb receptor activity suppresses colorectal cancer progression. *Nature* **435**, 1126–30 (2005).
- [121] Dallol, A., Morton, D., Maher, E. R. & Latif, F. Slit2 axon guidance molecule is frequently inactivated in colorectal cancer and suppresses growth of colorectal carcinoma cells. *Cancer Res* **63**, 1054–8 (2003).
- [122] Davalos, V. *et al.* Ephb4 and survival of colorectal cancer patients. *Cancer Res* **66**, 8943–8 (2006).
- [123] Dickinson, R. E. *et al.* Epigenetic inactivation of slit3 and slit1 genes in human cancers. *Br J Cancer* **91**, 2071–8 (2004).
- [124] Guo, D. L. *et al.* Reduced expression of ephb2 that parallels invasion and metastasis in colorectal tumours. *Carcinogenesis* **27**, 454–64 (2006).
- [125] Wang, J. *et al.* Downregulation of epha7 by hypermethylation in colorectal cancer. *Oncogene* **24**, 5637–47 (2005).
- [126] Li, V. S. *et al.* Frequent inactivation of axon guidance molecule rgma in human colon cancer through genetic and epigenetic mechanisms. *Gastroenterology* **137**, 176–87 (2009).
- [127] Zhao, Z. W. *et al.* Decreased expression of repulsive guidance molecule member a by dna methylation in colorectal cancer is related to tumor progression. *Oncol Rep* **27**, 1653–9 (2012).
- [128] Shi, Y. *et al.* Dragon (repulsive guidance molecule b, rgmb) is a novel gene that promotes colorectal cancer growth. *Oncotarget* (2015).
- [129] Li, J., Ye, L., Sanders, A. J. & Jiang, W. G. Repulsive guidance molecule b (rgmb) plays negative roles in breast cancer by coordinating bmp signaling. *J Cell Biochem* **113**, 2523–31 (2012).

- [130] Nohra, R. *et al.* Rgma and il21r show association with experimental inflammation and multiple sclerosis. *Genes Immun* **11**, 279–93 (2010).
- [131] Kujawa, S. G. & Liberman, M. C. Acceleration of age-related hearing loss by early noise exposure: evidence of a misspent youth. *J Neurosci* **26**, 2115–23 (2006).
- [132] Kujawa, S. G. & Liberman, M. C. Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. *J Neurosci* **29**, 14077–85 (2009).
- [133] McFadden, S. L., Ding, D., Jiang, H. & Salvi, R. J. Time course of efferent fiber and spiral ganglion cell degeneration following complete hair cell loss in the chinchilla. *Brain Res* **997**, 40–51 (2004).
- [134] Starr, A. *et al.* A dominantly inherited progressive deafness affecting distal auditory nerve and hair cells. *J Assoc Res Otolaryngol* **5**, 411–26 (2004).
- [135] White, J. A., Burgess, B. J., Hall, R. D. & Nadol, J. B. Pattern of degeneration of the spiral ganglion cell and its processes in the c57bl/6j mouse. *Hear Res* **141**, 12–8 (2000).
- [136] Camaschella, C., Roetto, A. & De Gobbi, M. Juvenile hemochromatosis. *Semin Hematol* **39**, 242–8 (2002).
- [137] De Gobbi, M. *et al.* Natural history of juvenile haemochromatosis. *Br J Haematol* **117**, 973–9 (2002).
- [138] Schaible, U. E. & Kaufmann, S. H. Iron and microbial infection. *Nat Rev Microbiol* **2**, 946–53 (2004).
- [139] Baraldi-Junkins, C. A., Beck, A. C. & Rothstein, G. Hematopoiesis and cytokines. relevance to cancer and aging. *Hematol Oncol Clin North Am* **14**, 45–61, viii (2000).
- [140] Bruunsgaard, H., Pedersen, M. & Pedersen, B. K. Aging and proinflammatory cytokines. *Curr Opin Hematol* **8**, 131–6 (2001).
- [141] Rajagopalan, S. *et al.* Neogenin mediates the action of repulsive guidance molecule. *Nat Cell Biol* **6**, 756–62 (2004).
- [142] Vielmetter, J., Kayyem, J. F., Roman, J. M. & Dreyer, W. J. Neogenin, an avian cell surface protein expressed during terminal neuronal differentiation, is closely related to the human tumor suppressor molecule deleted in colorectal cancer. *J Cell Biol* **127**, 2009–20 (1994).
- [143] Keino-Masu, K. *et al.* Deleted in colorectal cancer (dcc) encodes a netrin receptor. *Cell* **87**, 175–85 (1996).
- [144] Chan, S. S. *et al.* Unc-40, a c. elegans homolog of dcc (deleted in colorectal cancer), is required in motile cells responding to unc-6 netrin cues. *Cell* **87**, 187–95 (1996).

- [145] Geisbrecht, B. V., Dowd, K. A., Barfield, R. W., Longo, P. A. & Leahy, D. J. Netrin binds discrete subdomains of dcc and unc5 and mediates interactions between dcc and heparin. *J Biol Chem* **278**, 32561–8 (2003).
- [146] Finci, L. I. *et al.* The crystal structure of netrin-1 in complex with dcc reveals the bifunctionality of netrin-1 as a guidance cue. *Neuron* **83**, 839–49 (2014).
- [147] Xu, K. *et al.* Neural migration. structures of netrin-1 bound to two receptors provide insight into its axon guidance mechanism. *Science* **344**, 1275–9 (2014).
- [148] Yang, F., West, J., A. P. & Bjorkman, P. J. Crystal structure of a hemojuvelin-binding fragment of neogenin at 1.8a. *J Struct Biol* **174**, 239–44 (2011).
- [149] Stein, E., Zou, Y., Poo, M. & Tessier-Lavigne, M. Binding of dcc by netrin-1 to mediate axon guidance independent of adenosine a2b receptor activation. *Science* **291**, 1976–82 (2001).
- [150] Shekarabi, M. *et al.* Deleted in colorectal cancer binding netrin-1 mediates cell substrate adhesion and recruits cdc42, rac1, pak1, and n-wasp into an intracellular signaling complex that promotes growth cone expansion. *J Neurosci* **25**, 3132–41 (2005).
- [151] Li, X. *et al.* Netrin signal transduction and the guanine nucleotide exchange factor dock180 in attractive signaling. *Nat Neurosci* **11**, 28–35 (2008).
- [152] Liu, G. *et al.* Netrin requires focal adhesion kinase and src family kinases for axon outgrowth and attraction. *Nat Neurosci* **7**, 1222–32 (2004).
- [153] Conrad, S., Genth, H., Hofmann, F., Just, I. & Skutella, T. Neogenin-rgma signaling at the growth cone is bone morphogenetic protein-independent and involves rhoa, rock, and pkc. *J Biol Chem* **282**, 16423–33 (2007).
- [154] Endo, M. & Yamashita, T. Inactivation of ras by p120gap via focal adhesion kinase dephosphorylation mediates rgma-induced growth cone collapse. *J Neurosci* **29**, 6649–62 (2009).
- [155] Gad, J. M., Keeling, S. L., Wilks, A. F., Tan, S. S. & Cooper, H. M. The expression patterns of guidance receptors, dcc and neogenin, are spatially and temporally distinct throughout mouse embryogenesis. *Dev Biol* **192**, 258–73 (1997).
- [156] Keeling, S. L., Gad, J. M. & Cooper, H. M. Mouse neogenin, a dcc-like molecule, has four splice variants and is expressed widely in the adult mouse and during embryogenesis. *Oncogene* **15**, 691–700 (1997).
- [157] Meyerhardt, J. A., Look, A. T., Bigner, S. H. & Fearon, E. R. Identification and characterization of neogenin, a dcc-related gene. *Oncogene* **14**, 1129–36 (1997).

- [158] Rodriguez, A., Pan, P. & Parkkila, S. Expression studies of neogenin and its ligand hemojuvelin in mouse tissues. *J Histochem Cytochem* **55**, 85–96 (2007).
- [159] Konig, K. *et al.* The axonal guidance receptor neogenin promotes acute inflammation. *PLoS One* **7**, e32145 (2012).
- [160] Lee, D. H. *et al.* Neogenin inhibits hcv secretion and regulates bmp-induced hepcidin expression and iron homeostasis. *Blood* **115**, 3136–45 (2010).
- [161] Mawdsley, D. J. *et al.* The netrin receptor neogenin is required for neural tube formation and somitogenesis in zebrafish. *Dev Biol* **269**, 302–15 (2004).
- [162] Bragdon, B. *et al.* Bone morphogenetic proteins: a critical review. *Cell Signal* **23**, 609–20 (2011).
- [163] Nakase, T. & Yoshikawa, H. Potential roles of bone morphogenetic proteins (bmps) in skeletal repair and regeneration. *J Bone Miner Metab* **24**, 425–33 (2006).
- [164] Urist, M. R. & Strates, B. S. Bone morphogenetic protein. *J Dent Res* **50**, 1392–406 (1971).
- [165] Zhang, H. & Bradley, A. Mice deficient for bmp2 are nonviable and have defects in amnion/chorion and cardiac development. *Development* **122**, 2977–86 (1996).
- [166] Constam, D. B. & Robertson, E. J. Regulation of bone morphogenetic protein activity by pro domains and proprotein convertases. *J Cell Biol* **144**, 139–49 (1999).
- [167] Cui, Y., Jean, F., Thomas, G. & Christian, J. L. Bmp-4 is proteolytically activated by furin and/or pc6 during vertebrate embryonic development. *EMBO J* **17**, 4735–43 (1998).
- [168] Scheufler, C., Sebald, W. & Hulsmeyer, M. Crystal structure of human bone morphogenetic protein-2 at 2.7 a resolution. *J Mol Biol* **287**, 103–15 (1999).
- [169] Brown, M. A. *et al.* Crystal structure of bmp-9 and functional interactions with pro-region and receptors. *J Biol Chem* **280**, 25111–8 (2005).
- [170] Harrison, C. A., Al-Musawi, S. L. & Walton, K. L. Prodomains regulate the synthesis, extracellular localisation and activity of tgf-beta superfamily ligands. *Growth Factors* **29**, 174–86 (2011).
- [171] Little, S. C. & Mullins, M. C. Bone morphogenetic protein heterodimers assemble heteromeric type i receptor complexes to pattern the dorsoventral axis. *Nat Cell Biol* **11**, 637–43 (2009).

- [172] Mi, L. Z. *et al.* Structure of bone morphogenetic protein 9 procomplex. *Proc Natl Acad Sci U S A* **112**, 3710–5 (2015).
- [173] Sieber, C., Kopf, J., Hiepen, C. & Knaus, P. Recent advances in bmp receptor signaling. *Cytokine Growth Factor Rev* **20**, 343–55 (2009).
- [174] Shi, Y. & Massague, J. Mechanisms of tgf-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685–700 (2003).
- [175] Derynck, R. & Zhang, Y. E. Smad-dependent and smad-independent pathways in tgf-beta family signalling. *Nature* **425**, 577–84 (2003).
- [176] Gallea, S. *et al.* Activation of mitogen-activated protein kinase cascades is involved in regulation of bone morphogenetic protein-2-induced osteoblast differentiation in pluripotent c2c12 cells. *Bone* **28**, 491–8 (2001).
- [177] Guicheux, J. *et al.* Activation of p38 mitogen-activated protein kinase and c-jun-nh2-terminal kinase by bmp-2 and their implication in the stimulation of osteoblastic cell differentiation. *J Bone Miner Res* **18**, 2060–8 (2003).
- [178] Manning, B. D. & Cantley, L. C. Akt/pkb signaling: navigating downstream. *Cell* **129**, 1261–74 (2007).
- [179] Di Guglielmo, G. M., Le Roy, C., Goodfellow, A. F. & Wrana, J. L. Distinct endocytic pathways regulate tgf-beta receptor signalling and turnover. *Nat Cell Biol* **5**, 410–21 (2003).
- [180] Hartung, A. *et al.* Different routes of bone morphogenic protein (bmp) receptor endocytosis influence bmp signaling. *Mol Cell Biol* **26**, 7791–805 (2006).
- [181] Heining, E., Bhushan, R., Paarmann, P., Henis, Y. I. & Knaus, P. Spatial segregation of bmp/smud signaling affects osteoblast differentiation in c2c12 cells. *PLoS One* **6**, e25163 (2011).
- [182] Bragdon, B. *et al.* FRET reveals novel protein-receptor interaction of bone morphogenetic proteins receptors and adaptor protein 2 at the cell surface. *Biophys J* **97**, 1428–35 (2009).
- [183] Alborzinia, H. *et al.* Quantitative kinetics analysis of bmp2 uptake into cells and its modulation by bmp antagonists. *J Cell Sci* **126**, 117–27 (2013).
- [184] Shi, W. *et al.* Endofin acts as a smad anchor for receptor activation in bmp signaling. *J Cell Sci* **120**, 1216–24 (2007).
- [185] Callery, E. M. *et al.* Eps15r is required for bone morphogenetic protein signalling and differentially compartmentalizes with smad proteins. *Open Biol* **2**, 120060 (2012).

- [186] Gilboa, L. *et al.* Bone morphogenetic protein receptor complexes on the surface of live cells: a new oligomerization mode for serine/threonine kinase receptors. *Mol Biol Cell* **11**, 1023–35 (2000).
- [187] Nohe, A. *et al.* The mode of bone morphogenetic protein (bmp) receptor oligomerization determines different bmp-2 signaling pathways. *J Biol Chem* **277**, 5330–8 (2002).
- [188] Brazil, D. P., Church, R. H., Surae, S., Godson, C. & Martin, F. Bmp signalling: agony and antagonism in the family. *Trends Cell Biol* **25**, 249–64 (2015).
- [189] Xia, Y. *et al.* Repulsive guidance molecule rgma alters utilization of bone morphogenetic protein (bmp) type ii receptors by bmp2 and bmp4. *J Biol Chem* **282**, 18129–40 (2007).
- [190] Zhang, A. S., West, J., A. P., Wyman, A. E., Bjorkman, P. J. & Enns, C. A. Interaction of hemojuvelin with neogenin results in iron accumulation in human embryonic kidney 293 cells. *J Biol Chem* **280**, 33885–94 (2005).
- [191] Zhang, A. S., Yang, F., Wang, J., Tsukamoto, H. & Enns, C. A. Hemojuvelin-neogenin interaction is required for bone morphogenic protein-4-induced hepcidin expression. *J Biol Chem* **284**, 22580–9 (2009).
- [192] Xia, Y., Babitt, J. L., Sidis, Y., Chung, R. T. & Lin, H. Y. Hemojuvelin regulates hepcidin expression via a selective subset of bmp ligands and receptors independently of neogenin. *Blood* **111**, 5195–204 (2008).
- [193] Zhang, A. S. *et al.* Evidence that inhibition of hemojuvelin shedding in response to iron is mediated through neogenin. *J Biol Chem* **282**, 12547–56 (2007).
- [194] Yoon, B. S. & Lyons, K. M. Multiple functions of bmps in chondrogenesis. *J Cell Biochem* **93**, 93–103 (2004).
- [195] Yoon, B. S. *et al.* Bmpr1a and bmp1b have overlapping functions and are essential for chondrogenesis in vivo. *Proc Natl Acad Sci U S A* **102**, 5062–7 (2005).
- [196] Yoon, B. S. *et al.* Bmps regulate multiple aspects of growth-plate chondrogenesis through opposing actions on fgf pathways. *Development* **133**, 4667–78 (2006).
- [197] Zhou, Z. *et al.* Neogenin regulation of bmp-induced canonical smad signaling and endochondral bone formation. *Dev Cell* **19**, 90–102 (2010).
- [198] Tassew, N. G. *et al.* Modifying lipid rafts promotes regeneration and functional recovery. *Cell Rep* **8**, 1146–59 (2014).
- [199] Tian, C. *et al.* The neogenin/dcc homolog unc-40 promotes bmp signaling via the rgm protein drag-1 in *c. elegans*. *Development* **140**, 4070–80 (2013).

- [200] Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol Crystallogr* **62**, 1243–50 (2006).
- [201] Soding, J., Biegert, A. & Lupas, A. N. The hhpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244–8 (2005).
- [202] Yang, Z. R., Thomson, R., McNeil, P. & Esnouf, R. M. Ronn: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**, 3369–76 (2005).
- [203] Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. Signalp 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785–6 (2011).
- [204] UniProt, C. Uniprot: a hub for protein information. *Nucleic Acids Res* **43**, D204–12 (2015).
- [205] Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881–90 (1988).
- [206] von Einem, S., Schwarz, E. & Rudolph, R. A novel two-step renaturation procedure for efficient production of recombinant bmp-2. *Protein Expr Purif* **73**, 65–9 (2010).
- [207] Zhao, Y. *et al.* Automation of large scale transient protein expression in mammalian cells. *J Struct Biol* **175**, 209–15 (2011).
- [208] Chang, V. T. *et al.* Glycoprotein structural genomics: solving the glycosylation problem. *Structure* **15**, 267–73 (2007).
- [209] Davis, S. J. *et al.* Expression of soluble recombinant glycoproteins with predefined glycosylation: application to the crystallization of the t-cell glycoprotein cd2. *Protein Eng* **6**, 229–32 (1993).
- [210] Walter, T. S. *et al.* A procedure for setting up high-throughput nanolitre crystallization experiments. crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr D Biol Crystallogr* **61**, 651–7 (2005).
- [211] Mayo, C. J. *et al.* Benefits of automated crystallization plate tracking, imaging, and analysis. *Structure* **13**, 175–82 (2005).
- [212] Incardona, M. F. *et al.* Edna: a framework for plugin-based applications applied to x-ray experiment online data analysis. *J Synchrotron Radiat* **16**, 872–9 (2009).
- [213] Leslie, A. G. Integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **55**, 1696–702 (1999).

- [214] Miller, W. H. *A Treatise on Crystallography* (Cambridge Pitt Press for J. & J. J. Deighton, Trinity Street., Cambridge, UK, 1839).
- [215] Rupp, B. *Biomolecular Crystallography: Principles, Practice and Application to Structural Biology* (Garland Science, Taylor and Francis Group, LLC., USA, 2010).
- [216] Laue, M. v. Röntgenstrahlinterferenzen. *Physikalische Zeitschrift* **14**, 1075–1079 (1913).
- [217] Laue, M. v. Zur optik der raumgitter. *Physikalische Zeitschrift* **14**, 1040–1041 (1913).
- [218] Argand, R. *Essai sur une manière de représenter les quantités imaginaires dans les constructions géométriques* (Gauthier-Villars, Paris, France, 1874), second edn.
- [219] Bragg, W. L. The diffraction of short electromagnetic waves by a crystal. *Proc. Cambridge Phil. Soc.* **17**, 43–57 (1913).
- [220] Ewald, P. Introduction to the dynamical theory of x-ray diffraction. *Acta Crystallographica Section A* **25**, 103–108 (1969).
- [221] Sauter, N. K., Grosse-Kunstleve, R. W. & Adams, P. D. Robust indexing for automatic data collection. *J Appl Crystallogr* **37**, 399–409 (2004).
- [222] Zhang, Z., Sauter, N. K., van den Bedem, H., Snell, G. & Deacon, A. M. Automated diffraction image analysis and spot searching for high-throughput crystal screening. *Journal of Applied Crystallography* **39**, 112–119 (2006).
- [223] Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62**, 48–57 (2006).
- [224] Kabsch, W. Automatic indexing of rotation diffraction patterns. *Journal of Applied Crystallography* **21**, 67–72 (1988).
- [225] Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography* **26**, 795–800 (1993).
- [226] Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125–32 (2010).
- [227] Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **276**, 307–326 (1997).
- [228] Leslie, A. G. W. & Powell, H. R. *Processing diffraction data with Mosflm*, vol. 245, 41–51 (NATO Science series, Springer, Dordrecht, 2007).
- [229] Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72–82 (2006).

- [230] Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**, 186–190 (2010).
- [231] Arndt, U. W., Crowther, R. A. & Mallett, J. F. A computer-linked cathode-ray tube microdensitometer for x-ray crystallography. *J Sci Instrum* **1**, 510–6 (1968).
- [232] Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–3 (2012).
- [233] Diederichs, K. & Karplus, P. A. Improved r-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* **4**, 269–75 (1997).
- [234] Weiss, M. Global indicators of x-ray data quality. *Journal of Applied Crystallography* **34**, 130–135 (2001).
- [235] Diederichs, K. & Karplus, P. A. Better models by discarding data? *Acta Crystallogr D Biol Crystallogr* **69**, 1215–22 (2013).
- [236] Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204–14 (2013).
- [237] Patterson, A. L. A direct method for the determination of the components of interatomic distances in crystals. *Zeitschrift für Kristallographie* **A90**, 517–542 (1935).
- [238] McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658–674 (2007).
- [239] Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 470–8 (2010).
- [240] Bell, C. H. *et al.* Structure of the repulsive guidance molecule (rgm)-neogenin signaling hub. *Science* **341**, 77–80 (2013).
- [241] Kirsch, T., Sebald, W. & Dreyer, M. K. Crystal structure of the bmp-2-bria ectodomain complex. *Nat Struct Biol* **7**, 492–6 (2000).
- [242] Kleywegt, G. J. Validation of protein crystal structures. *Acta Crystallogr D Biol Crystallogr* **56**, 249–65 (2000).
- [243] Cowtan, K. The buccaneer software for automated model building. 1. tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**, 1002–11 (2006).
- [244] Terwilliger, T. C. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D Biol Crystallogr* **59**, 38–44 (2003).
- [245] Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**, 95–9 (1963).

- [246] Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–32 (2004).
- [247] Bricogne, G. *et al.* Buster version 2.8.0 (2011).
- [248] Adams, P. D. *et al.* Phenix: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–21 (2010).
- [249] Brunger, A. T. Free r value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–5 (1992).
- [250] Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated ncs and target-structure restraints in buster. *Acta Crystallogr D Biol Crystallogr* **68**, 368–80 (2012).
- [251] Davis, I. W. *et al.* Molprobity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375–83 (2007).
- [252] Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037–41 (2001).
- [253] The pymol molecular graphics system, version 1.7.2.
- [254] Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774–97 (2007).
- [255] van der Merwe, P. A. *Protein-Ligand Interactions: A Practical Approach*, vol. 1, 137–170 (Oxford University Press, Oxford, UK, 2001).
- [256] Johnsson, B., Löfås, S. & Lindquist, G. Immobilization of proteins to a carboxymethyl-dextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. *Analytical Biochemistry* **198**, 268–77 (1991).
- [257] Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (saxs) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics* **40**, 191–285 (2007).
- [258] Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477–481 (2013).
- [259] Pelikan, M., Hura, G. L. & Hammel, M. Structure and flexibility within proteins as identified through small angle x-ray scattering. *General physiology and biophysics* **28**, 174–189 (2009).
- [260] Pernot, P. *et al.* Upgraded esrf bm29 beamline for saxs on macromolecules in solution. *Journal of Synchrotron Radiation* **20**, 660–664 (2013).

- [261] Petoukhov, M. V. *et al.* New developments in the atsas program package for small-angle scattering data analysis. *Journal of Applied Crystallography* **45**, 342–350 (2012).
- [262] Leaver-Fay, A. *et al.* *Chapter nineteen - Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules*, vol. Volume 487, 545–574 (Academic Press, 2011).
- [263] Eswar, N. *et al.* Tools for comparative protein structure modeling and analysis. *Nucleic Acids Research* **31**, 3375–3380 (2003).
- [264] Weinkam, P., Pons, J. & Sali, A. Structure-based model of allostery predicts coupling between distant sites. *Proceedings of the National Academy of Sciences* **109**, 4875–4880 (2012).
- [265] Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate saxs profile computation and its assessment by contrast variation experiments. *Biophysical Journal* **105**, 962–974 (2013).
- [266] Guttman, M., Weinkam, P., Sali, A. & Lee, K. K. All-atom ensemble modeling to analyze small-angle x-ray scattering of glycosylated proteins. *Structure* **21**, 321–31 (2013).
- [267] Korchynskyi, O. & ten Dijke, P. Identification and functional characterization of distinct critically important bone morphogenetic protein-specific response elements in the id1 promoter. *J Biol Chem* **277**, 4883–91 (2002).
- [268] Owen, D. M. *et al.* Palm imaging and cluster analysis of protein heterogeneity at the cell surface. *J Biophotonics* **3**, 446–54 (2010).
- [269] Rossy, J., Owen, D. M., Williamson, D. J., Yang, Z. & Gaus, K. Conformational states of the kinase lck regulate clustering in early t cell signaling. *Nat Immunol* **14**, 82–9 (2013).
- [270] Han, R., Li, Z., Fan, Y. & Jiang, Y. Recent advances in super-resolution fluorescence imaging and its applications in biology. *J Genet Genomics* **40**, 583–95 (2013).
- [271] Padilla-Parra, S. & Tramier, M. FRET microscopy in the living cell: different approaches, strengths and weaknesses. *Bioessays* **34**, 369–76 (2012).
- [272] Perry, G. L. W. Sppack: spatial point pattern analysis in excel using visual basic for applications (vba). *Environmental Modelling & Software* **19**, 559–569 (2004).
- [273] Team, R. C. R: A language and environment for statistical computing. (2013).
- [274] Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. Nih image to imagej: 25 years of image analysis. *Nat Methods* **9**, 671–5 (2012).

- [275] Williamson, D. J. *et al.* Pre-existing clusters of the adaptor lat do not participate in early t cell signaling events. *Nat Immunol* **12**, 655–62 (2011).
- [276] Wilkinson, M. H. F. Optimizing edge detectors for robust automatic threshold selection: Coping with edge curvature and noise. *Graphical Models and Image Processing* **60**, 385–401 (1998).
- [277] Padilla-Parra, S., Auduge, N., Coppey-Moisán, M. & Tramier, M. Quantitative fret analysis by fast acquisition time domain flim at high spatial resolution in living cells. *Biophys J* **95**, 2976–88 (2008).
- [278] Brown, P. H., Balbo, A. & Schuck, P. Using prior knowledge in the determination of macromolecular size-distributions by analytical ultracentrifugation. *Biomacromolecules* **8**, 2011–24 (2007).
- [279] Schuck, P. & Rossmanith, P. Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers* **54**, 328–41 (2000).
- [280] Yang, F., West, J., A. P., Allendorph, G. P., Choe, S. & Bjorkman, P. J. Neogenin interacts with hemojuvelin through its two membrane-proximal fibronectin type iii domains. *Biochemistry* **47**, 4237–45 (2008).
- [281] Landon. Cleavage at aspartyl-prolyl bonds. *Methods Enzymol* **47**, 145–9 (1977).
- [282] Lidell, M. E. & Hansson, G. C. Cleavage in the gdph sequence of the c-terminal cysteine-rich part of the human muc5ac mucin. *Biochem J* **399**, 121–9 (2006).
- [283] Ambrosio, A. L. *et al.* Crossveinless-2 is a bmp feedback inhibitor that binds chordin/bmp to regulate xenopus embryonic patterning. *Dev Cell* **15**, 248–60 (2008).
- [284] Zhang, A. S. *et al.* Neogenin-mediated hemojuvelin shedding occurs after hemojuvelin traffics to the plasma membrane. *J Biol Chem* **283**, 17494–502 (2008).
- [285] Daopin, S., Piez, K. A., Ogawa, Y. & Davies, D. R. Crystal structure of transforming growth factor-beta 2: an unusual fold for the superfamily. *Science* **257**, 369–73 (1992).
- [286] Griffith, D. L., Keck, P. C., Sampath, T. K., Rueger, D. C. & Carlson, W. D. Three-dimensional structure of recombinant human osteogenic protein 1: structural paradigm for the transforming growth factor beta superfamily. *Proc Natl Acad Sci U S A* **93**, 878–83 (1996).
- [287] Sampath, T. K. & Reddi, A. H. Dissociative extraction and reconstitution of extracellular matrix components involved in local bone differentiation. *Proc Natl Acad Sci U S A* **78**, 7599–603 (1981).

- [288] Wozney, J. M. *et al.* Novel regulators of bone formation: molecular clones and activities. *Science* **242**, 1528–34 (1988).
- [289] Ruppert, R., Hoffmann, E. & Sebald, W. Human bone morphogenetic protein 2 contains a heparin-binding site which modifies its biological activity. *Eur J Biochem* **237**, 295–302 (1996).
- [290] Attisano, L., Wrana, J. L., Montalvo, E. & Massague, J. Activation of signalling by the activin receptor complex. *Mol Cell Biol* **16**, 1066–73 (1996).
- [291] Knaus, P. & Sebald, W. Cooperativity of binding epitopes and receptor chains in the bmp/tgfbeta superfamily. *Biol Chem* **382**, 1189–95 (2001).
- [292] Koenig, B. B. *et al.* Characterization and cloning of a receptor for bmp-2 and bmp-4 from nih 3t3 cells. *Mol Cell Biol* **14**, 5961–74 (1994).
- [293] Wrana, J. L., Attisano, L., Wieser, R., Ventura, F. & Massague, J. Mechanism of activation of the tgfbeta receptor. *Nature* **370**, 341–7 (1994).
- [294] Heinecke, K. *et al.* Receptor oligomerization and beyond: a case study in bone morphogenetic proteins. *BMC Biol* **7**, 59 (2009).
- [295] Mueller, T. D. & Nickel, J. Promiscuity and specificity in bmp receptor activation. *FEBS Lett* **586**, 1846–59 (2012).
- [296] Greenwald, J. *et al.* The bmp7/actrii extracellular domain complex provides new insights into the cooperative nature of receptor assembly. *Mol Cell* **11**, 605–17 (2003).
- [297] Greenwald, J., Fischer, W. H., Vale, W. W. & Choe, S. Three-finger toxin fold for the extracellular ligand-binding domain of the type ii activin receptor serine kinase. *Nat Struct Biol* **6**, 18–22 (1999).
- [298] Allendorph, G. P., Vale, W. W. & Choe, S. Structure of the ternary signaling complex of a tgfbeta superfamily member. *Proc Natl Acad Sci U S A* **103**, 7643–8 (2006).
- [299] Lebrun, J. J. & Vale, W. W. Activin and inhibin have antagonistic effects on ligand-dependent heteromerization of the type i and type ii activin receptors and human erythroid differentiation. *Mol Cell Biol* **17**, 1682–91 (1997).
- [300] Groppe, J. *et al.* Cooperative assembly of tgfbeta superfamily signaling complexes is mediated by two disparate mechanisms and distinct modes of receptor binding. *Mol Cell* **29**, 157–68 (2008).
- [301] Thompson, T. B., Woodruff, T. K. & Jardetzky, T. S. Structures of an actriib:activin a complex reveal a novel binding mode for tgfbeta ligand:receptor interactions. *EMBO J* **22**, 1555–66 (2003).

- [302] Wu, Q., Sun, C. C., Lin, H. Y. & Babitt, J. L. Repulsive guidance molecule (rgm) family proteins exhibit differential binding kinetics for bone morphogenetic proteins (bmps). *PLoS One* **7**, e46307 (2012).
- [303] Castonguay, R. *et al.* Soluble endoglin specifically binds bone morphogenetic proteins 9 and 10 via its orphan domain, inhibits blood vessel formation, and suppresses tumor growth. *J Biol Chem* **286**, 30034–46 (2011).
- [304] Herrera, B. & Inman, G. J. A rapid and sensitive bioassay for the simultaneous measurement of multiple bone morphogenetic proteins. identification and quantification of bmp4, bmp6 and bmp9 in bovine and human serum. *BMC Cell Biol* **10**, 20 (2009).
- [305] Barthe, P. *et al.* Solution structure of human p8mtcp1, a cysteine-rich protein encoded by the mtcp1 oncogene, reveals a new alpha-helical assembly motif. *J Mol Biol* **274**, 801–15 (1997).
- [306] Obita, T. *et al.* Structural basis for selective recognition of escrt-iii by the aaa atpase vps4. *Nature* **449**, 735–9 (2007).
- [307] Hammel, M. *et al.* Characterization of ehp, a secreted complement inhibitory protein from staphylococcus aureus. *J Biol Chem* **282**, 30051–61 (2007).
- [308] Camus, L. M. & Lambert, L. A. Molecular evolution of hemojuvelin and the repulsive guidance molecule family. *J Mol Evol* **65**, 68–81 (2007).
- [309] Liao, S. M., Du, Q. S., Meng, J. Z., Pang, Z. W. & Huang, R. B. The multiple roles of histidine in protein interactions. *Chem Cent J* **7**, 44 (2013).
- [310] Weber, D. *et al.* A silent h-bond can be mutationally activated for high-affinity interaction of bmp-2 and activin type iib receptor. *BMC Struct Biol* **7**, 6 (2007).
- [311] Kanomata, K., Kokabu, S., Nojima, J., Fukuda, T. & Katagiri, T. Dragon, a gpi-anchored membrane protein, inhibits bmp signaling in c2c12 myoblasts. *Genes Cells* **14**, 695–702 (2009).
- [312] Leung, J. C., Tsang, A. W., Chan, D. T. & Lai, K. N. Absence of cd89, polymeric immunoglobulin receptor, and asialoglycoprotein receptor on human mesangial cells. *J Am Soc Nephrol* **11**, 241–9 (2000).
- [313] Stirling, C. M. *et al.* Characterization of the porcine neonatal fc receptor—potential use for trans-epithelial protein delivery. *Immunology* **114**, 542–53 (2005).
- [314] Silvestri, L., Pagani, A. & Camaschella, C. Furin-mediated release of soluble hemojuvelin: a new link between hypoxia and iron homeostasis. *Blood* **111**, 924–31 (2008).

- [315] Monteiro, R. M. *et al.* Real time monitoring of bmp smads transcriptional activity during mouse development. *Genesis* **46**, 335–46 (2008).
- [316] Sountoulidis, A. *et al.* Activation of the canonical bone morphogenetic protein (bmp) pathway during lung morphogenesis and adult lung tissue repair. *PLoS One* **7**, e41460 (2012).
- [317] Pagani, A., Silvestri, L., Nai, A. & Camaschella, C. Hemojuvelin n-terminal mutants reach the plasma membrane but do not activate the hepcidin response. *Haematologica* **93**, 1466–72 (2008).
- [318] Vieira, A. V., Lamaze, C. & Schmid, S. L. Control of egf receptor signaling by clathrin-mediated endocytosis. *Science* **274**, 2086–9 (1996).
- [319] Le Roy, C. & Wrana, J. L. Clathrin- and non-clathrin-mediated endocytic regulation of cell signalling. *Nat Rev Mol Cell Biol* **6**, 112–26 (2005).
- [320] Healey, E. G. *et al.* Repulsive guidance molecule is a structural bridge between neogenin and bone morphogenetic protein. *Nat Struct Mol Biol* **22**, 458–65 (2015).
- [321] Hagiwara, M. *et al.* Neogenin, a receptor for bone morphogenetic proteins. *J Biol Chem* **286**, 5157–65 (2011).
- [322] Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J Mol Biol* **234**, 946–50 (1993).
- [323] Bork, P. The modular architecture of a new family of growth regulators related to connective tissue growth factor. *FEBS Lett* **327**, 125–30 (1993).
- [324] Gum, J., J. R., Hicks, J. W., Toribara, N. W., Siddiki, B. & Kim, Y. S. Molecular cloning of human intestinal mucin (muc2) cDNA. identification of the amino terminus and overall sequence similarity to prepro-von willebrand factor. *J Biol Chem* **269**, 2440–6 (1994).
- [325] Meiniel, O. & Meiniel, A. The complex multidomain organization of spondin protein is highly conserved in mammals. *Brain Res Rev* **53**, 321–7 (2007).
- [326] Ota, T. *et al.* Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet* **36**, 40–5 (2004).
- [327] Hu, H. Cell-surface heparan sulfate is involved in the repulsive guidance activities of slit2 protein. *Nat Neurosci* **4**, 695–701 (2001).
- [328] Hussain, S. A. *et al.* A molecular mechanism for the heparan sulfate dependence of slit-robo signaling. *J Biol Chem* **281**, 39693–8 (2006).
- [329] Inatani, M., Irie, F., Plump, A. S., Tessier-Lavigne, M. & Yamaguchi, Y. Mammalian brain morphogenesis and midline axon guidance require heparan sulfate. *Science* **302**, 1044–6 (2003).

- [330] Garcia-Garcia, M. J. & Anderson, K. V. Essential role of glycosaminoglycans in fgf signaling during mouse gastrulation. *Cell* **114**, 727–37 (2003).
- [331] Jackson, S. M. *et al.* dally, a drosophila glypican, controls cellular responses to the tgf-beta-related morphogen, dpp. *Development* **124**, 4113–20 (1997).
- [332] Xiao, Y. *et al.* Rgmb is a novel binding partner for pd-l2 and its engagement with pd-l2 promotes respiratory tolerance. *J Exp Med* **211**, 943–59 (2014).
- [333] Guha, A., Sriram, V., Krishnan, K. S. & Mayor, S. Shibire mutations reveal distinct dynamin-independent and -dependent endocytic pathways in primary cultures of drosophila hemocytes. *J Cell Sci* **116**, 3373–86 (2003).
- [334] Kalia, M. *et al.* Arf6-independent gpi-anchored protein-enriched early endosomal compartments fuse with sorting endosomes via a rab5/phosphatidylinositol-3'-kinase-dependent machinery. *Mol Biol Cell* **17**, 3689–704 (2006).
- [335] Ricci, V. *et al.* High cell sensitivity to helicobacter pylori vaca toxin depends on a gpi-anchored protein and is not blocked by inhibition of the clathrin-mediated pathway of endocytosis. *Mol Biol Cell* **11**, 3897–909 (2000).
- [336] Skretting, G., Torgersen, M. L., van Deurs, B. & Sandvig, K. Endocytic mechanisms responsible for uptake of gpi-linked diphtheria toxin receptor. *J Cell Sci* **112** (Pt 22), 3899–909 (1999).
- [337] Brown, D. Gpi-anchored proteins and detergent-resistant membrane domains. *Braz J Med Biol Res* **27**, 309–15 (1994).
- [338] Edidin, M. Lipids on the frontier: a century of cell-membrane bilayers. *Nat Rev Mol Cell Biol* **4**, 414–8 (2003).
- [339] Friedrichson, T. & Kurzchalia, T. V. Microdomains of gpi-anchored proteins in living cells revealed by crosslinking. *Nature* **394**, 802–5 (1998).
- [340] Kenworthy, A. K., Petranova, N. & Edidin, M. High-resolution fret microscopy of cholera toxin b-subunit and gpi-anchored proteins in cell plasma membranes. *Mol Biol Cell* **11**, 1645–55 (2000).
- [341] Simons, K. & Ikonen, E. Functional rafts in cell membranes. *Nature* **387**, 569–72 (1997).
- [342] Varma, R. & Mayor, S. Gpi-anchored proteins are organized in submicron domains at the cell surface. *Nature* **394**, 798–801 (1998).
- [343] Schlessinger, J. & Ullrich, A. Growth factor signaling by receptor tyrosine kinases. *Neuron* **9**, 383–91 (1992).

- [344] van der Geer, P., Hunter, T. & Lindberg, R. A. Receptor protein-tyrosine kinases and their signal transduction pathways. *Annu Rev Cell Biol* **10**, 251–337 (1994).
- [345] Haigler, H. T., McKanna, J. A. & Cohen, S. Direct visualization of the binding and internalization of a ferritin conjugate of epidermal growth factor in human carcinoma cells a-431. *J Cell Biol* **81**, 382–95 (1979).
- [346] Damke, H., Baba, T., Warnock, D. E. & Schmid, S. L. Induction of mutant dynamin specifically blocks endocytic coated vesicle formation. *J Cell Biol* **127**, 915–34 (1994).
- [347] Di Guglielmo, G. M., Baass, P. C., Ou, W. J., Posner, B. I. & Bergeron, J. J. Compartmentalization of shc, grb2 and msos, and hyperphosphorylation of raf-1 by egf but not insulin in liver parenchyma. *EMBO J* **13**, 4269–77 (1994).
- [348] Miyazono, K., ten Dijke, P. & Heldin, C. H. Tgf-beta signaling by smad proteins. *Adv Immunol* **75**, 115–57 (2000).
- [349] Tsukazaki, T., Chiang, T. A., Davison, A. F., Attisano, L. & Wrana, J. L. Sara, a fyve domain protein that recruits smad2 to the tgfbeta receptor. *Cell* **95**, 779–91 (1998).
- [350] Paratcha, G., Ledda, F. & Ibanez, C. F. The neural cell adhesion molecule ncam is an alternative signaling receptor for gdnf family ligands. *Cell* **113**, 867–79 (2003).
- [351] Tansey, M. G., Baloh, R. H., Milbrandt, J. & Johnson, J., E. M. Gfralpha-mediated localization of ret to lipid rafts is required for effective downstream signaling, differentiation, and neuronal survival. *Neuron* **25**, 611–23 (2000).
- [352] Topalian, S. L. *et al.* Safety, activity, and immune correlates of anti-pd-1 antibody in cancer. *N Engl J Med* **366**, 2443–54 (2012).
- [353] Cunha, S. I. *et al.* Genetic and pharmacological targeting of activin receptor-like kinase 1 impairs tumor growth and angiogenesis. *J Exp Med* **207**, 85–100 (2010).
- [354] Hu-Lowe, D. D. *et al.* Targeting activin receptor-like kinase 1 inhibits angiogenesis and tumorigenesis through a mechanism of action complementary to anti-vegf therapies. *Cancer Res* **71**, 1362–73 (2011).
- [355] Baud’huin, M. *et al.* A soluble bone morphogenetic protein type ia receptor increases bone mass and bone strength. *Proc Natl Acad Sci U S A* **109**, 12207–12 (2012).
- [356] Zhou, X. *et al.* Reversal of cancer cachexia and muscle wasting by actriib antagonism leads to prolonged survival. *Cell* **142**, 531–43 (2010).