Structural studies of chemotaxis in prokaryotes and higher eukaryotes

Christian H. Bell

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



Wolfson College
Department of Biochemistry
Division of Structural Biology
University of Oxford

Trinity 2011

As the creeper that girdles the tree-trunk,

the Law runneth forward and back;

For the strength of the Pack is the Wolf,

and the strength of the Wolf is the Pack.

The Jungle Book, Rudyard Kipling

Structural studies of chemotaxis in prokaryotes and higher eukaryotes

Christian H. Bell

A thesis submitted in partial fulfilment of the requirements for
the degree of *Doctor of Philosophy*
Wolfson College, Department of Biochemistry and Division of Structural Biology
University of Oxford, Trinity Term 2011

Abstract

Chemotaxis, the ability of a cell to change its motility as a response to environmental stimuli, belongs to the most important signal transduction mechanisms in life. Evolution has created a plethora of different molecular mechanisms to enable cells to react appropriately to extracellular changes. In this thesis, chemotactic signal transduction pathways in bacteria, in particular two component signalling in *R. sphaeroides*, and eukaryotes, in particular human axon guidance, are studied by means of X-ray crystallography complemented with biophysical, biochemical and cellular studies.

Two-component signal transduction in bacteria is one of the most abundant signalling pathways in biology. Here the determinants for specificity for a crucial sensor kinase-response regulator interaction are presented and used to design novel, synthetic two-component pairs. The activation mechanism of response regulators has been extensively studied and a model for activation that crucially involves a Thr and a Tyr residue has been widely accepted in the field. In this thesis, two structures of a response regulator from *R. sphaeroides* are presented and together with biophysical and cellular assays suggest a novel paradigm for response regulator activation.

Axon guidance is an essential process in human development and relies crucially on chemotaxis. Two signalling pathways, the plexin-semaphorin and the Rgm-Neogenin pathway are studied extensively in this work. Structures of the intracellular region of Plexin-B1 provide an elegant mechanism explaining how ligand binding events on the extracellular and intracellular side can be integrated into a single signalling output. The study of RgmB in complex with its receptor Neogenin provides the first structural insight into the important family of repulsive guidance molecules and explains their role in the disease juvenile hemachromatosis.

In summary, this work provides insights into a plethora of chemotactic pathways and extends our current knowledge of these important mechanisms significantly.

Acknowledgement

Firstly, I would like to thank my supervisors Judy Armitage, Yvonne Jones, Christian Siebold and Dave Stuart for their encouragement and support during the four years of my DPhil studies. Strubi is a fantastic place to work in and I am grateful to them for giving me the opportunity to study here. Especially in combination with the equipment available at the Department of Biochemistry, there was never a time when experiments could not be done due to lack of technical possibilities or know-how. I also want to thank all of my supervisors for their patience and trust which enabled me to work independently on several completely different projects.

I would also like to thank everyone in Strubi and the Armitage Lab for helping me in the wetlab and making my four years in Oxford so very much enjoyable. I would like to acknowledge Nicola Abrescia, Benjamin Bishop, Amber Clayton, Alexandra Flanagan, Stephen Graham, Bert Janssen, Margaret Jones, Weixian Lu, Tomas Malinauskas, Ross Robinson and Elena Seiradake at Strubi and Mostyn Brown, Nicolas Delalez, Kathryn Scott, David Wilkinson in the Armitage Lab for their help in learning new techniques, great discussions and weekend support. Special thanks go to Radu Aricescu and Steve Porter for their practical help and the many useful discussions. I would also like to acknowledge Karl Harlos and Tom Walter for their invaluable help in setting up crystals and preparing them for data collection.

Finally I want to thank Lena and my family for their constant love and support during my time in Oxford.

Table of contents

Abstract 1

Acknowledgement 2

Table of contents 3

Chapter 1 Introduction 8

1.1 Chemotaxis 8

1.2 Bacterial chemotaxis 9

1.2.1 Chemotaxis in *E. coli* 10

1.2.2 Chemotaxis in *R. sphaeroides* 13

1.2.2.1 *R. sphaeroides* has two chemosensory clusters 15

1.3 Eukaryotic chemotaxis 19

1.3.1 Plexin-semaphorin signalling cues 20

1.3.1.1 Small GTPases in plexin-semaphorin signalling 24

1.3.1.2 Structural basis for extracellular plexin-semaphorin signalling 28

1.3.2 Repulsive guidance molecules 31

1.3.2.1 RgmA 32

1.3.2.2 RgmB 33

1.3.2.3 RgmC 34

1.3.2.4 Neogenin signalling 35

1.3.2.5 BMP signalling 37

1.4 Aims 40

Chapter 2 Materials and Methods 42

2.1 Molecular Biology 42

2.1.1 cDNA and expression vectors 42

2.1.2 Construct design and PCR 43

2.1.3 Restriction digests, ligations and transformations 43
2.1.4 Site-directed mutagenesis 44
2.2 Protein expression and purification 45
2.2.1 Protein expression and purification in *E. coli* 45
2.2.2 Protein expression and purification from insect cells 47
2.2.3 Protein expression and purification in mammalian cells 48
2.3 Protein crystallization 50
2.4 Basic principles of crystallography 51
2.4.1 Description of crystals and Miller indices 51
2.4.2 Crystal symmetry 52
2.4.3 X-rays 52
2.4.4 The scattering process 53
2.4.5 Bragg's law 55
2.4.6 The reciprocal space 56
2.4.7 The structure factor 57
2.4.8 Calculation of the electron density 58
2.4.9 Friedel's law 59
2.4.10 Anomalous scattering 59
2.4.11 The temperature factor 60
2.5 Data collection and processing 60
2.6 Structure solution and refinement 61
2.6.1 Phasing using Se-SAD 61
2.6.2 Phasing using molecular replacement 62
2.6.3 Refinement 63
2.7 Surface Plasmon Resonance 65
2.7.1 Sample preparation 65
2.7.2 Experimental setup 66
2.8 Phosphotransfer assay 67
2.9 Swarm Plate assay 68

2.10 Thermofluor assay 68
2.11 COS-7-cell collapse assay 69
2.12 BMP response assay....70
2.13 Analytical ultracentrifugation....70
2.14 Figures....70

Chapter 3 Rewiring chemotaxis from the structure of a histidine phosphotransfer domain in complex with its cognate response regulator 73

3.1 Summary 73
3.2 Introduction 74
3.3 Expression and purification of $CheY_6$ and $CheA_3P1$ 76
3.4 Crystallization and structure determination of the $CheY_6$-$CheA_3P1$ complex 77
3.5 Structure of the $CheA_3P1$-$CheY_6$ complex 81
3.6 Structure of phosphorylated $CheA_3P1$ in complex with $CheY_6$ Asp56Asn,Ser83Ala .84
3.7 The interface between $CheA_3P1$ and $CheY_6$ 86
3.8 Biophysical and functional characterization of the interface....91
3.9 Rewiring chemotaxis 95
3.10 Discussion 98

Chapter 4 Crystal structures of $CheY_3$ - a new paradigm for response regulator activation? 103

4.1 Summary 103
4.2 Introduction 104
4.3 Expression and purification of $CheY_3$....106
4.4 Crystallization and structure determination....108
4.5 Crystal structure of $CheY_3$....112
4.6 Biophysical and functional characterization of $CheY_3$ 118
4.6 Discussion....124

Chapter 5 A dual binding mode for RhoGTPases in plexin signalling 128

5.1 Summary 128

5.2 Introduction 128

5.3 Expression and purification of Plexin-B1 and Rac1 131

5.4 Crystallization and structure determination of the Plexin-B1-Rac1* complex 133

5.5 Structure of the intracellular domain of Plexin-B1 in complex with constitutively active Rac1 136

5.6 Three-fold arrangement of the Plexin-$B1_{cyto}$-Rac1 unit 146

5.7 The importance of the N-terminal segment of the Plexin-B1 cytoplasmic region 154

5.8 Discussion 158

Chapter 6 Structural studies of the Repulsive Guidance Molecule B in complex with its receptor Neogenin 164

6.1 Summary 164

6.2 Introduction 164

6.3 Expression and purification of RgmB and Neogenin 167

6.4 Crystallization and structure determination of the Neo_{56}-RgmB complex 171

6.5 The structure of RgmB reveals a novel protein fold 174

6.6 The interaction between RgmB and its receptor Neogenin 182

6.7 Structural and functional implications for the role of RgmB in BMP signalling 197

6.8 Discussion 200

Chapter 7 Conclusion 205

Appendix A Expression vectors 208

Appendix B Abbreviations 212

Appendix C Publications 216

References 217

Chapter 1

Introduction

Chapter 1 Introduction

1.1 Chemotaxis

The ability to respond to environmental stimuli is one of the most important determinants of life. Evolution has created a vast array of signalling mechanisms that enable cells to react in multiple ways to environmental changes. One of these mechanisms is chemotaxis, the capability of biasing motility in response to extracellular stimuli. Chemotactic signalling can be found in bacteria, archaea and eukaryotes and represents a crucial part of signal transduction. However, despite huge efforts, even its simplest form, two-component signalling in bacteria, is not yet perfectly understood. The signalling pathways are distinct and mostly very specific, yet the underlying principle is conserved among all aspects of life. Initially, chemoattractants or -repellents bind to a transmembrane or intracellular receptor. This binding event leads to conformational changes within the receptor which then allows the binding and regulation of intracellular proteins (e.g. prokaryotic response regulators or eukaryotic small GTPases). These so-called molecular switches trigger further signalling events ultimately leading to movement of the cell either towards (in case of an attractant, positive chemotaxis) or away (repellent, negative chemotaxis) from the chemical stimulus. Chemotaxis plays a key role in the virulence of many pathogenic bacteria (e.g. *Helicobacter pylori* [1] and *Pseudomonas aeruginosa* [2]) and is also involved in the pathogenesis of inflammatory diseases, such as infections, allergies, and atherosclerosis, as well as in cancer [3,4].

This thesis will investigate three different chemotactic signalling pathways. Firstly, two-component signalling in *R. sphaeroides* and in particular the role of specificity therein will be examined. In addition, the molecular switch of a response regulator from inactive to active conformation will be discussed. As examples of eukayotic chemotaxis, the plexin-semaphorin and the repulsive guidance molecule signalling pathways will be studied extensively.

1.2 Bacterial chemotaxis

Bacteria need to be able to respond and adapt to changing environmental conditions to react to, for example, scarcity of nutrients or the accumulation of toxins and chemotaxis allows them to do so. Motility in most bacteria is generated by rotating helical filaments called flagella at speeds of up to 1,000 hertz that drive the cell forwards. They are powered by the flagellar motor, a multi-protein complex that spans the inner and outer membrane and is powered by an electrochemical proton or, in some cases, Na^+ gradient across the cytoplasmic membrane [5]. Chemotaxis regulates the rotation of the flagellar motor thereby biasing motility. The signal transduction pathway underlying chemotaxis starts with a chemoreceptor sensing a change in the environmental conditions. This signal is then propagated through a histidine protein kinase and a response regulator to the flagellar motor. The interactions between histidine kinases and their cognate response regulators play vital parts in many aspects of the bacterial lifecycle and they are also referred to as two-component signalling pathways.

Since most bacteria are too small to sense concentration gradients along their cell length, they alter their motility by temporal sensing. This results in a biased random walk whereby smooth swimming periods are alternated with brief direction changes. The mechanism that bacteria use for these direction changes are not uniform across different species. *Escherichia coli* has a flagellar motor that can rotate in both directions. Counterclockwise rotation leads to smooth swimming, whereas clockwise rotation leads to disassembly of the flagellar bundle and to random tumbling. *Rhodobacter sphaeroides* has a single flagellum that only rotates in one direction. A change of direction is achieved by stopping the motor. Brownian motion then reorientates the cell body and allows for a different direction once rotation is resumed. The frequency of direction changes is controlled by the chemotaxis pathway. When bacteria swim towards favourable conditions, a change in direction is suppressed, whereas such a change is increased when they swim in an unfavourable direction.

1.2.1 Chemotaxis in *E. coli*

The best studied chemotactic system is that of *E. coli*. The change in environmental conditions is sensed through certain key molecules called chemoeffectors. These can either act as chemoattractants or repellents. Chemoeffectors are detected by chemoreceptors that are dimeric transmembrane proteins comprising of a ligand-binding periplasmic domain and a cytoplasmic signalling domain [6]. Upon binding of the chemoeffector a piston like movement of the transmembrane domain is induced that transduces the signal from the periplasm to the cytosol. Thousands of chemoreceptors are clustered into large arrays

usually found at the cell poles [7]. These arrays are highly structured and the chemoreceptors are organized into units of trimers of dimers. These units interact with the histidine protein kinase CheA and the linker protein CheW to form ternary signalling complexes. The clustering of many receptors and the allosteric interactions within these amplify the signal and allow bacteria to respond to very small changes in chemoeffector concentration [8].

The dimeric histidine protein kinase CheA can phosphorylate the response regulators CheY and CheB in one of the best understood two-component pathways. CheA is a multidomain protein consisting of five separate structural domains connected by short linker regions. The P1 domain is the histidine phosphotransfer domain that transfers the phosphoryl group onto the aspartate residue of a cognate response regulator. The P2 domain provides additional binding surface for the response regulators. P3 is responsible for dimerization of the CheA whereas P4 is the kinase domain that binds ATP and phosphorylates a His residue (His48) on P1. The final domain, P5, binds CheW and the chemoreceptors and regulates the kinase activity. In unfavourable conditions the chemoreceptors promote the autophosphorylation of CheA [9]. The phosphorylated P1 domain then transfers the phosphoryl onto the response regulator CheY and CheB (Fig. 1.1). Phosphorylated CheY (CheY-P) then leaves the polar clusters and diffuses to the flagellar motor where it interacts with FliM and FliN of the flagellar switch complex to promote a switch in the rotational direction [10]. In order to sense chemoeffector gradients continuously the signal termination, dephosphorylation of CheY-P, has to be rapid. Since response regulators only have very slow auto-dephosphorylation rates, this is catalysed in *E. coli* by the specific phosphatase CheZ (Fig. 1.1).

CheB plays a crucial role in adaptation. This allows bacteria to compare current and previous conditions by resetting the chemotactic pathway to a time-averaged chemoeffector concentration [11]. On a molecular level, adaptation occurs through the reversible methylation of the cytoplasmic domains of the chemoreceptor. Methylation results in an increased ability to activate CheA. In *E. coli*, CheR methylates specific Glu residues on the chemorecepor whereas CheB removes these posttranslational modifications (Fig. 1.1). Phosphorylation of CheB results in a 100-fold higher activity [12]. In a non-favourable environment, autophosphorylation of CheA is increased, leading to higher levels of CheY-P and CheB-P. CheY-P results in tumbling of the cell whereas, albeit on a slower timescale, CheB-P demethylates the chemoreceptors thus decreasing their ability to stimulate CheA autophosphorylation. This negative feedback loop allows the bacterium to reset the tumble bias to a pre-stimulus level, thereby adapting to the environment.

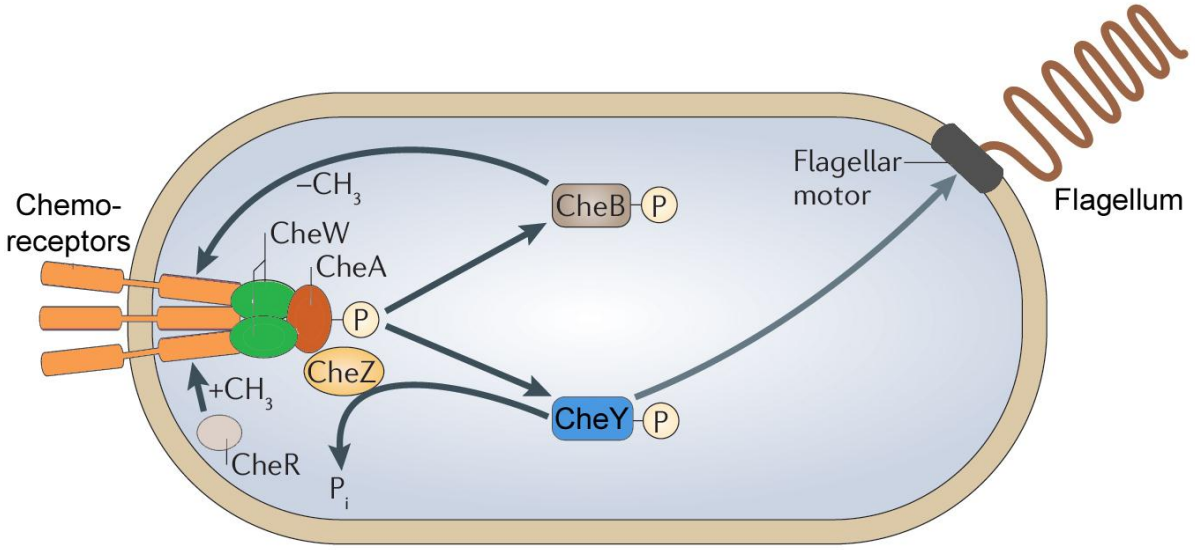


Figure 1.1: The chemotactic pathway of *E. coli*. Chemoeffectors are sensed by the chemoreceptors. These can promote CheA autophosphorylation, which in turn leads to phosphotransfer to the response regulators CheY and CheB. CheY-P alters the direction of rotation of the flagellar motor. CheB-P and CheR are responsible for adaptation modifying the chemoreceptors. CheZ is a specific phosphatase involved in signal termination (adapted from Porter et al. [237]).

E. coli has a simple chemotactic system with only one kinase CheA, and only two response regulators, one being responsible for switching of the flagellar motor (CheY), the other one for adaptation (CheB). All bacterial species share the same overall architecture of the signalling pathway, however, most of them are much more complex with additional chemotaxis proteins, cytoplasmic chemoreceptors and different adaptation or signal termination strategies [13,14]. One of the best understood of these complex networks is found in *Rhodobacter sphaeroides*.

1.2.2 Chemotaxis in *R. sphaeroides*

R. sphaeroides is a model organism to study photosynthesis, bioenergetics, cell biology and signal transduction. It can grow under anaerobic phototrophic conditions as well as aerobically and in the absence of light. It is metabolically diverse and is able to fix nitrogen. In accordance with this diversity, its chemoeffectors span a wide range of stimuli including light, oxygen, amino acids, sugars and organic acids such as succinate or propionate. Unlike *E. coli*, *R. sphaeroides* uses energy taxis, which describes movement towards an environment that is optimal for metabolic activity rather than having the highest attractant concentration. Thus, transport and metabolism of the chemoattractants are needed for sensing [15].

The chemotactic signalling pathway is more complex than in *E. coli* and contains 13 chemoreceptors (four of which are cytoplasmic), four CheA-like proteins, four CheW homologues, six CheY homologues, two CheB-like proteins and three CheR homologues. In addition, *R. sphaeroides* also has two sets of flagellar genes, each encoding functional

flagella [16]. The reason for these two separate flagellar systems is unclear. Under laboratory conditions, the first set of flagellar genes, *fla1*, is used exclusively and a *fla1* deletion leads to a loss of motility. However, all sequenced *R. sphaeroides* strains have intact *fla2* gene sets, suggesting that these genes are expressed under growth conditions that differ from the ones used in the laboratory [17].

Each set of flagellar genes has a characteristic set of chemotaxis proteins regulating them. The chemotactic proteins are expressed from three operons, *che*Op1, *che*Op2 and *che*Op3 (Fig. 1.2). Gene products of *che*Op1 control the rotation of the Fla2 flagellum, whereas *che*Op2 and *che*Op3 are responsible for regulating Fla1. The expression of the latter operons is controlled by two separate σ-factors, σ^{28} and a σ^{54} homologue, respectively. This additional level of regulation might allow variation of expression of these two operons in response to different environmental conditions, thereby enabling *R. sphaeroides* to tune its chemotactic response.

The Fla1 flagellar motor rotates in one direction only and motility is regulated by stopping the motor. The frequency and duration of the stops is controlled by the chemotaxis pathway. Stopping of the motor is achieved by a powerful molecular brake with the binding of the phosphorylated response regulator resulting in locking of the flagellum [18]. It is unclear why different species have different numbers of flagella and why some bacteria employ switching rather than stopping the flagellar motor. One possibility could be that certain swimming patterns are beneficial and more efficient in certain viscosities or in environments with specific chemoeffector gradients [19].

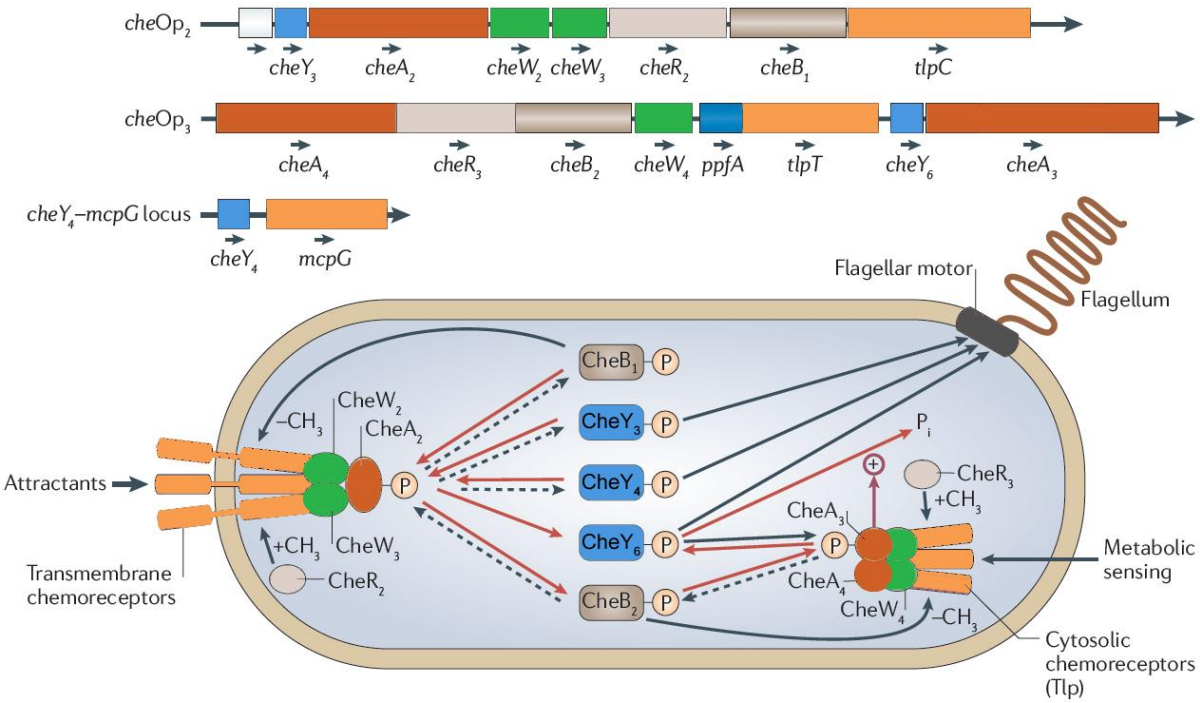


Figure 1.2: The chemotactic pathway of *R. sphaeroides*. The genes encoding the chemotactic proteins that regulate Fla1 are shown on top. They are clustered into two groups, chemotaxis operon 2 (*che*Op2) and chemotaxis operon 3 (*che*Op3), and an additional locus encoding for $CheY_4$ and the chemoreceptor McpG. There are two distinct signalling clusters, one located at the cell pole and one in the cytoplasm. The former senses attractant concentrations in the periplasm whereas the latter is responsible for detecting the metabolic state of the cell. Red arrows show the phosphorylation reactions that allow $CheY_6$ to act as phosphate sink for the other response regulators. Dashed arrows mark the phosphorylation reactions that connect the polar with the cytoplasmic cluster via the $CheB_2$-mediated phosphorelay.(adapted from Porter et al. [237]).

1.2.2.1 *R. sphaeroides* has two chemosensory clusters

Unlike other bacteria with much simpler chemotactic pathways, *R. sphaeroides* localises its chemosensory proteins to two different chemosensory clusters. One is located at the cell pole, similar to the chemotactic cluster in *E. coli*, whereas the other resides in the cytoplasm. Both clusters are necessary for chemotaxis and neither of the two clusters alone can support chemotactic responses. The cytoplasmic cluster is characterized by including soluble chemoreceptors. They lack the transmembrane region of their polar cluster

equivalents and allow sensing of cytoplasmic rather than periplasmic concentration of chemoeffectors. The exact chemical identity of chemoeffectors for the cytoplasmic cluster is unknown, however, it is clear that many attractants have to be transported and metabolised for chemotaxis to occur [20,21]. The intracellular cluster provides a mechanism for energy taxis in which optimal concentrations of metabolites rather than high concentrations of extracellular chemoattractants are guiding the cell into different environments, tuning the response according to current energy requirements.

The architecture of the polar cluster is very similar to the one found in *E. coli* with the majority of its components encoded by *che*Op2. The formation of the cluster is crucially dependant on $CheA_2$, $CheW_2$, $CheW_3$ and the chemoreceptors [22]. $CheA_2$ autophosphorylation is triggered by changes in the periplasmic concentration of chemorepellents. Once phosphorylated, the histidine kinase can then phosphorylate all of the *R. sphaeroides* response regulators (Fig. 1.3a).

The cytoplasmic cluster comprises of the cytoplasmic chemoreceptors and the chemotactic proteins encoded on *che*Op3. The former along with $CheW_4$ and either $CheA_3$ or $CheA_4$ are essential for cluster formation [22]. Similar to its polar counterpart, the cytoplasmic cluster contains over 1,000 chemoreceptors and associated chemosensory proteins. It is not known how this vast number of proteins can form a quarternary structure in particular in the absence of a membrane scaffold. Cells typically have a single cytoplasmic cluster located at mid-cell. Upon cell division this cluster is split into two and each half migrates to the one-quarter and three-quarter cell position. This ensures that each daughter cell inherits one

cluster. The separation is mediated by PpfA, which is a homologue of the ParA family responsible for ATP-driven chromosome and plasmid partitioning.

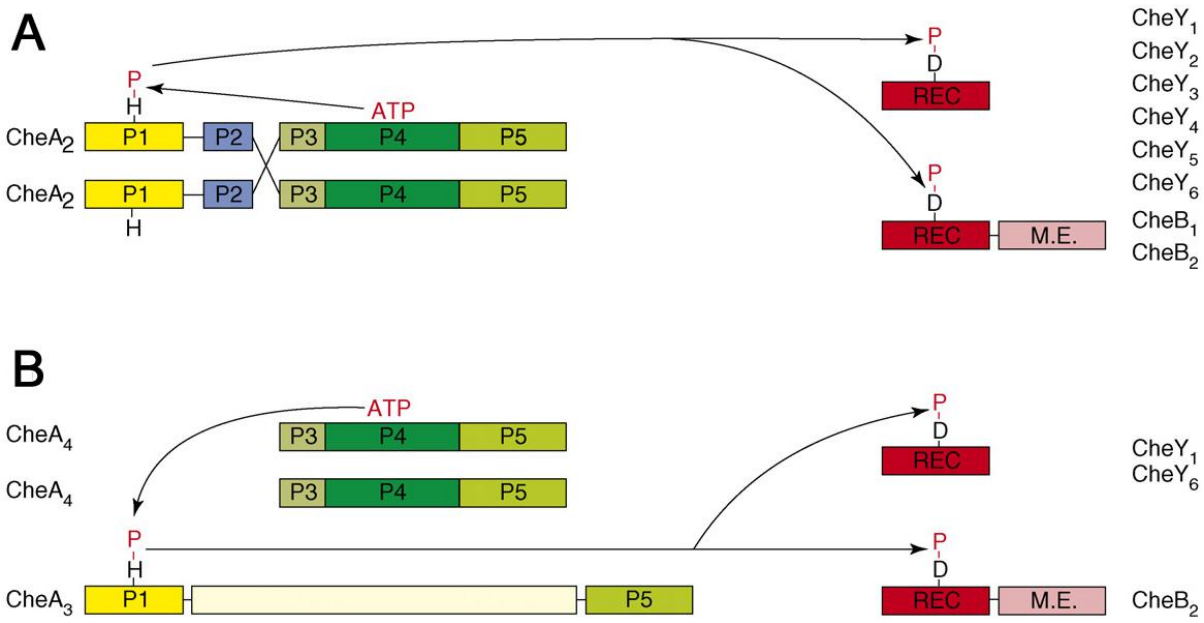


Fig. 1.3: Summary of the chemotaxis phosphorylation reactions showing the domain structures of the proteins involved. (a) A $CheA_2$ homodimer autophosphorylates on a histidine residue within the P1 domain using ATP as phosphodonor. $CheA_2$-P phosphotransfers to all of the *R. sphaeroides* CheY and CheB proteins. (c) A $CheA_4$ homodimer phosphorylates the P1 domain of $CheA_3$ using ATP as phosphodonor. $CheA_3$-P phosphotransfers to $CheY_1$, $CheY_6$ and $CheB_2$. REC denotes the receiver domains found in CheY and CheB proteins; M.E. denotes the methylesterase domain found in CheB proteins (adapted from Porter et al. [138]).

The histidine kinases in the cytoplasmic cluster $CheA_4$ and $CheA_3$ form an unusual complex that is crucial for chemotactic responses (Fig. 1.3b). Compared to *E. coli* CheA, $CheA_3$ is truncated at the N-terminus and only contains the P3-P5 domain. $CheA_4$ consists of a P1 and a P5 domain separated by a 794-residue long region with no structural homology. This region has been shown to have specific phosphatase activity towards $CheY_6$-P [23]. Upon

stimulation the kinase domain of $CheA_4$ phosphorylates the P1 domain of $CheA_3$ on His51. $CheA_3$ can then phosphorylate $CheY_1$, $CheY_6$ or $CheB_2$ [24]. Since $CheY_1$ is part of the Fla2 set and not expressed under laboratory conditions, the cytoplasmic cluster only directly regulates two response regulators, $CheY_6$ and $CheB_2$. The receptor binding domains of $CheA_3$ and $CheA_4$ are different and thus their different regulation could play an important role in this pathway.

R. sphaeroides has six different CheY proteins of which only three are expressed under laboratory conditions, $CheY_3$, $CheY_4$ and $CheY_6$. $CheY_6$ and either $CheY_3$ or $CheY_4$ are essential for chemotaxis. $CheY_3$ and $CheY_4$ share a high sequence identity and are redundant for each other. Only $CheY_6$-P is able to stop the flagellar motor and thus bias the motility of the cell. It can do so even in the absence of $CheY_3$ or $CheY_4$, however, without the latter the response is non-chemotactic [25]. All three CheY-homologues can bind to the flagellar motor but the effects of binding of $CheY_3$-P and $CheY_4$-P on the flagellar motor are not known [26].

Fast signal termination is necessary for efficient chemotaxis. However, all response regulators have auto-dephosphorylation rates that are too slow to allow for a quick response and in *R. sphaeroides* only one phosphatase domain, located on $CheA_3$ and specific for $CheY_6$-P has been identified so far. Some bacteria use so called phosphate sinks instead of phosphatases to terminate a signal. In *R. sphaeroides* mathematical modelling suggests a similar mechanism. It involves reversed phospho-transfer from the response regulators, $CheY_3$-P and $CheY_4$-P onto the histidine kinase and from there onto the phosphate sink $CheY_6$. Since $CheY_6$ can be dephosphorylated rapidly and is a target of both the cytoplasmic and polar cluster it can act as phosphate sink for all other response regulators [27] (Fig. 1.2).

In addition, the two chemotactic clusters are also connected by a phosphorelay system involving $CheB_2$. $CheA_3$ from the cytoplasmic cluster can phosphorylates $CheB_2$ and $CheB_2$-P is able to phosphorylate $CheA_2$ in the polar cluster. Thus, this phosphorelay allows $CheA_3$ to indirectly phosphorylate every response regulator in the cell and therefore affect their activation levels (Fig. 1.2).

1.3 Eukaryotic chemotaxis

In eukaryotes, chemotaxis guides the migration of cells, such as neurons, leukocytes, stem cells, and simple amoeba, along chemical gradients. Unlike bacteria, chemotactic sensing is not just temporal but also spatial, meaning that cells can sense gradients along their cell length [28]. This is achieved by different receptor occupancies resulting in a chemotactic response only towards the direction with a high concentration of chemoeffectors. Movement of the cells is mediated by alterations of the eukaryotic cytoskeleton. This motility depends on both myosin-mediated contraction at the rear of the cell and actin polymerization at the front. The actin cytoskeleton is crucially regulated by small GTPases of the Rho-family and indeed disruption of some of the individual Rho proteins or their putative regulators and effectors in *Dictyostelium* results in impaired chemotaxis [29]. However, the exact mechanisms by which receptors regulate Rho-proteins and ultimately the actin cytoskeleton remain elusive.

Chemotaxis plays a vital part in axon guidance where different occupancies of chemoreceptors polarize growth cones to move either towards an attractant or away from a repellent. Unlike in bacteria, chemoeffectors in eukaryotes are not limited to small

molecules but instead often compromise of secreted or membrane-bound proteins. Several ligand-receptor pairs have been described to be important for axon guidance. Among them are Slit-Robo [30], Semaphorin-Plexin [31], Ephrin-Eph [32], Netrin-DCC [33] and Rgm-Neogenin [34]. The molecular mechanisms by which these signalling pairs initiate signal transduction cascades leading to reorganisation of the cytoskeleton and a bias of motility are mostly unclear despite some recent structural insights [35-39].

1.3.1 Plexin-semaphorin signalling cues

Semaphorins are a family of secreted or membrane-anchored glycoproteins that signal through the transmembrane receptor Plexin. Upon semaphorin binding plexins can interact with small GTPases of the Rho and Ras family and regulate their activity [31].

Plexin-semaphorin signalling plays an important role in angiogenesis, skeletal development, tumor progression and immune responses, however, it is most well studied in the context of repulsive axonal guidance where it plays a crucial role in initiating remodelling of the actin cytoskeleton [40,41].

Plexins are receptors for semaphorins and can be grouped into four classes based on their overall homology. Sequence analyses indicate an N-terminal sema-domain, followed by a combination of three PSI domains and six IPT domains (Ig domain shared by plexins and transcription factors). The intracellular domain of plexins shares homology with the GTPase activating (GAP) domain of p120 RasGAP and is highly conserved between different plexins. The GAP domain forms an integral structural unit that has an additional domain inserted

into one of its loop regions [43,44]. This additional domain can bind to small GTPases of the Rho-family and is thus called the Rho-GTPase binding domain (RBD).

Most plexin–semaphorin interactions are mediated through the sema domains of both proteins, except for class 3 semaphorins, which require neuropilins as essential co-receptors to signal through class A plexins [45,46]. Neuropilins are transmembrane proteins whose extracellular region can be segmented into five domains, two CUB (complement C1r/C1s, Uegf, Bmp1) domains followed by two Coagulation factor 5/8 type domains and a MAM (derived from meprin, A-5 protein, and receptor protein-tyrosine phosphatase mu) domain. The intracellular region lacks any enzymatic activity.

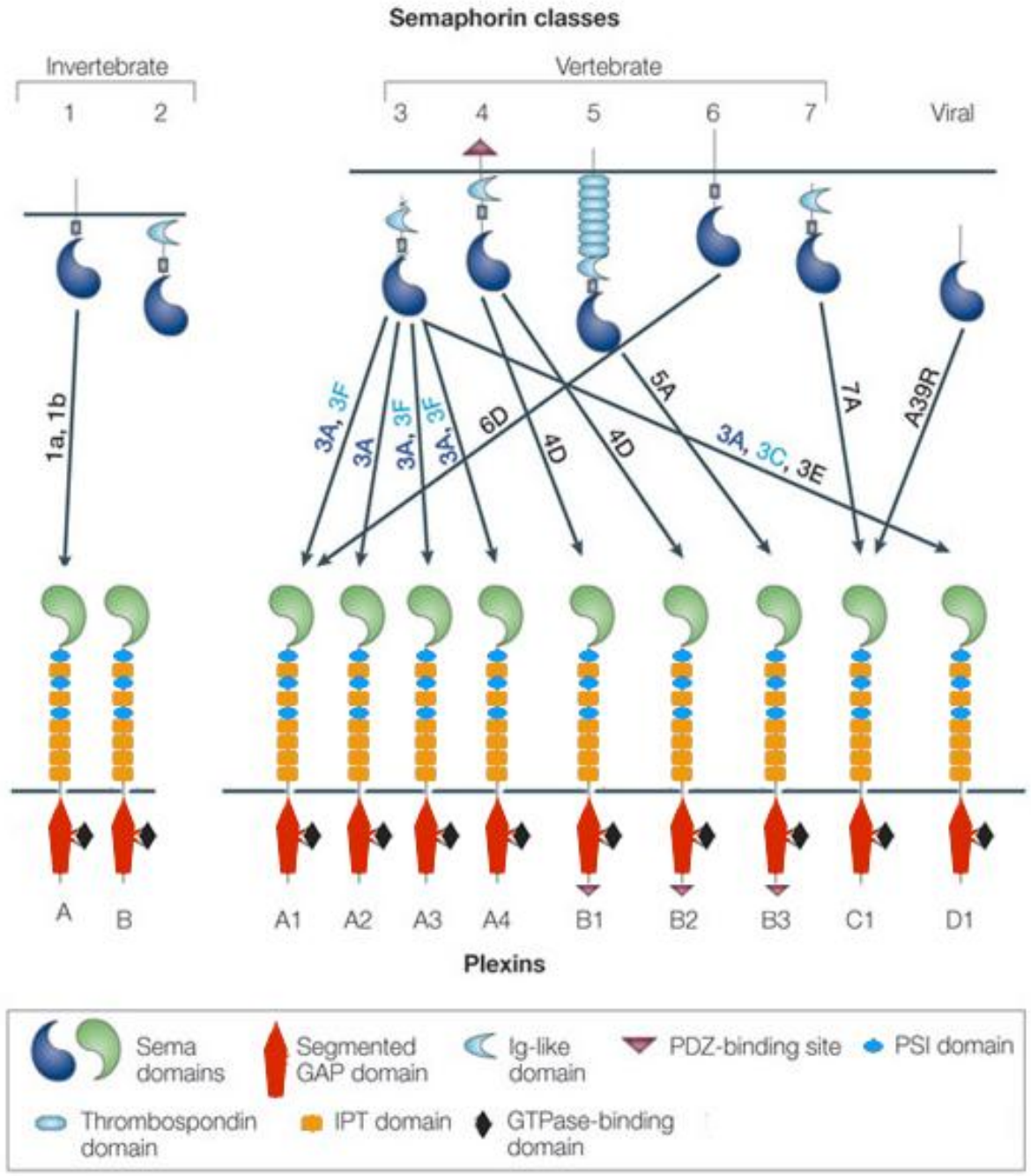


Figure 1.3: Overview over different semaphorin classes and their plexin receptors. Both semaphorins and plexins have the characteristic sema domain. Semaphorins have additional domains including PSI (plexin, semaphorin and integrin) domains, immunoglobulin (Ig)-like domains, thrombospondin domains and PDZ-domain-binding sites. In addition to the sema domain plexins have PSI domains, IPT (Ig-like, plexins and transcription factors) domains, a GTPase-binding domain and a segment GAP (GTPase-activating protein) domain. Some plexins also have PDZ-domain binding sites. Arrows indicate binding interactions detected between semaphorins and plexins. Labels on the arrows indicate which specific semaphorin interacts with which plexin. Blue labels indicate the necessity for neuropilin-1 (dark blue) or neuropilin-2 (light blue) as co-receptors. (adapted from Kruger et al. [31])

There are eight classes of semaphorins, grouped based on their sequence and structural similarity and over 20 different semaphorins have been described so far (Fig. 1.3). Their central structural feature is a seven-bladed β-propeller called the sema-domain that can also be found in plexins as well as Met and Ron receptors [42]. All semaphorins have a PSI (plexin, semaphorin, integrin) domain C-terminal to the sema domain followed by a variety of additional domains depending on their class.

The conserved intracellular plexin GAP domain is crucial for plexin function. Mutation of its catalytic Arg residues leads to loss of function in Plexin-A1 and a loss of the Plexin-B1 mediated collapse response in COS-7 cells [47,48]. In Plexin-B1 the targets for the plexin GAP activity are two members of the Ras family, R-Ras and M-Ras [49,50]. The extracellular semaphorin-plexin interaction alone is not sufficient for this GAP activity to occur [49]. In addition, a small GTPase of the Rho-family has to be bound to the RBD of the intracellular domain. This intriguing signalling mechanism that integrates both extra- and intracellular ligand binding events into a single signalling output is not fully understood yet.

Plexins have also been shown to interact with GTPase-exchange factors (GEFs), such as the Rac-GEF FARP2 [51] and PDZ-Rho-GEF [52] thereby promoting the activation of RhoGTPases. In contrast, the cytoplasmic domain of Plexin-A1 and Plexin-B1 can also interact with p190 Rho-GAP [53] catalyzing RhoGTPase inactivation. As a result, the regulation of RhoGTPases might vary in different cells which is in agreement with the observation that plexins have been found to trigger multiple intracellular pathways, sometimes leading to opposing functional effects [54].

1.3.1.1 Small GTPases in plexin-semaphorin signalling

Plexin-semaphorin signalling crucially relies on small GTPases of the Rho- and Ras family. These small GTPases act as molecular switches and exist in two different conformations, the GTP-bound, active state and the GDP-bound, inactive state. There are more than 100 small GTPases that have been described in different, eukaryotic organisms.

Ras family members are key regulators in a variety of pathways, most importantly cell proliferation. Ras is the most common human proto-oncogene since single point mutations can lead to a permanently active Ras which is deficient in intrinsic or stimulated GTP hydrolysis [55]. This results in uncontrolled proliferation which potentially leads to tumor genesis [56]. Members of the Rho family play a major role in the regulation of the actin cytoskeleton. They participate in a plethora of biochemical pathways including morphogenesis, phagocytosis, axon outgrowth, cell-cell adhesion, cell polarity, cell cycle regulation and cell motility [57].

All small GTPase family members have a conserved fold consisting of a six stranded β-sheet surrounded by five α-helices [58,59] (Fig. 1.5). Members of the Rho-family have an additional insert between the β_5-strand and the α_4-helix [60]. Small GTPases have three characteristic, structural elements: the phosphate binding loop, the switch I and the switch II region (Fig. 1.5). In the GTP-bound state the switch regions form two hydrogen bonds with the γ-phosphate of GTP. After GTP hydrolysis this phosphate group is released allowing the two switch regions to relax into the GDP-specific conformation. The conformational change can be described as loaded spring mechanism [58]. The phosphate binding loop and other amino

acids participating in nucleotide binding are highly conserved. Mutations in this region have dramatic effects on the intrinsic hydrolysis and nucleotide dissociation rate.

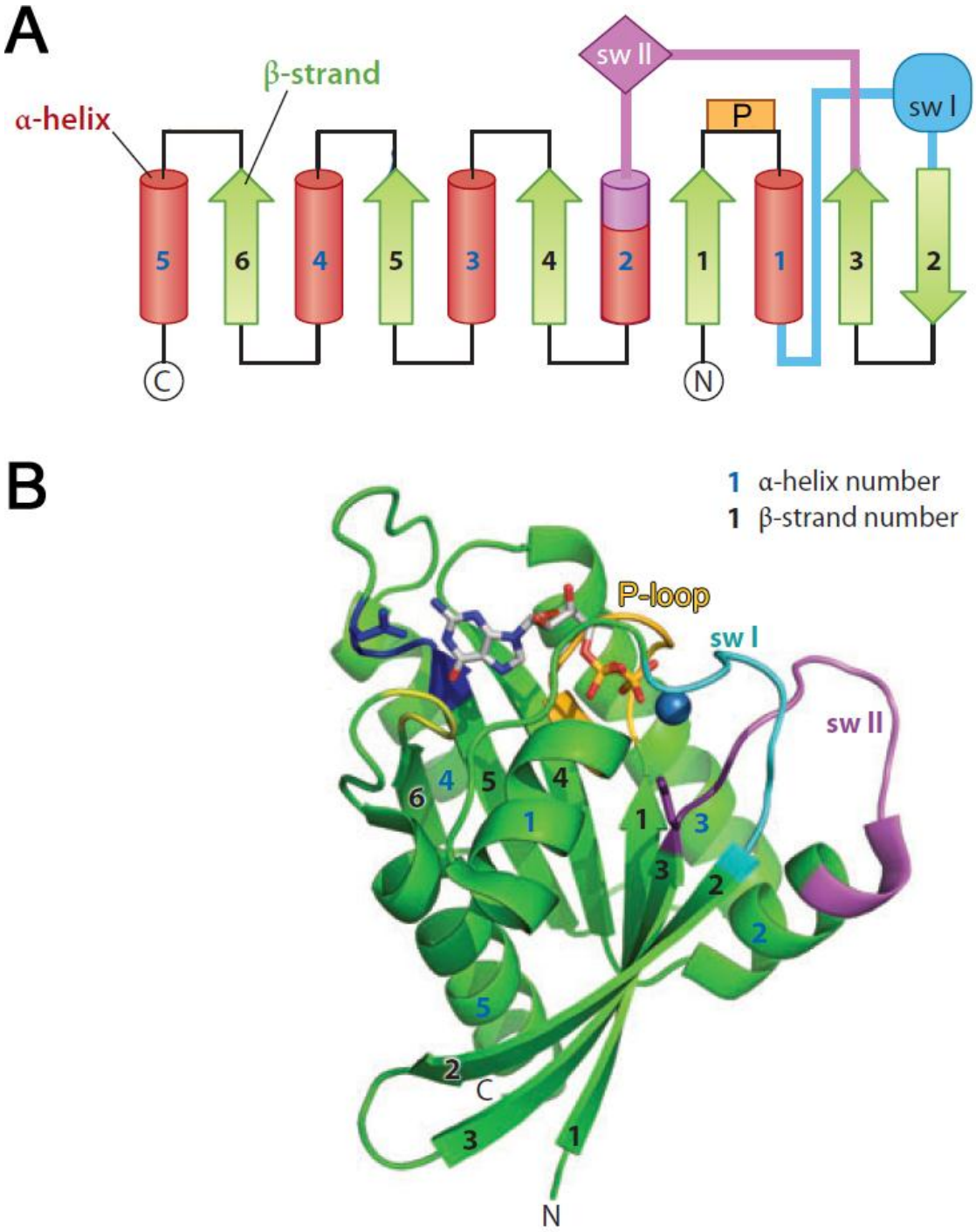


Figure 1.5: General features of small GTPases. (A) Topology diagram of the G domain with β-strands in green, α-helices in red, and N and C termini as indicated; (b) Ras•GppNHp•Mg2+ (PDB-ID: 5P21) structure as the prototype G domain. The conserved sequence motifs including the switch regions and the P-loop are labelled (adapted from Wittinghofer et al. [59])

In order to be able to respond to signals as quickly as possible, the conformational change from GDP-bound inactive to GTP-bound, active conformation and vice versa has to happen very fast. However, the intrinsic nucleotide exchange rate as well as the GTP hydrolysis rate are slow and thus not suitable for most biological signal transduction processes which require complete conformational change within minutes after GTP loading and GTP hydrolysis, respectively. Therefore these reactions are further regulated by three groups of proteins:

1. Guanine nucleotide exchange factors (GEFs)
2. GTPase activating proteins (GAPs)
3. Guanine nucleotide dissociation inhibitors (GDIs)

GEFs are responsible for the activation of small GTPases by catalyzing the dissociation of GDP [61] (Fig. 1.4). The binding of GTP and GDP, respectively, underlies a concentration dependent equilibrium reaction since the affinity of the small GTPase for both these nucleotides is approximately equal. Nucleotide binding of a nucleotide-free small GTPase is strongly shifted towards GTP binding, because the cellular concentration of GTP is about 100-fold higher than that of GDP.

Upon activation by GEFs, small GTPases are then able to interact with effectors and further downstream targets [62] (Fig. 1.4). The interaction is dependent on the GTP-bound active state of the GTPase since activation-induced conformational changes in the switch regions are essential for molecular recognition of the effector.

The signalling cycle is terminated by GAPs that catalyze the intrinsic hydrolysis by five orders of magnitude [58,63,64]. GAPs stabilize the transition state structure for example by introducing

an Arg residue into the active site of the GTPase [65,66]. This so-called Arg-finger interacts with the nucleotide's β- and γ-phosphate thereby shielding their charge and facilitating the nucleophilic attack of a water molecule.

GDIs are characteristic for the regulation of members of the Rho and Rab family [58,59]. They cause an internal translocation of membrane-bound GTPases. Small GTPases are lipid-modified at their C-terminus thereby enabling them to be anchored to the membrane. GDIs bind to this hydrophobic tail and shield it from the aqueous cytosolic environment thus forming a cytoplasmic pool of lipid-modified proteins. This then allows recycling of these proteins between different parts of the cellular membrane. In addition, the translocation of the small GTPase from the membrane to the cytosol prevents interaction with membrane bound GEFs and thus GTPase activation.

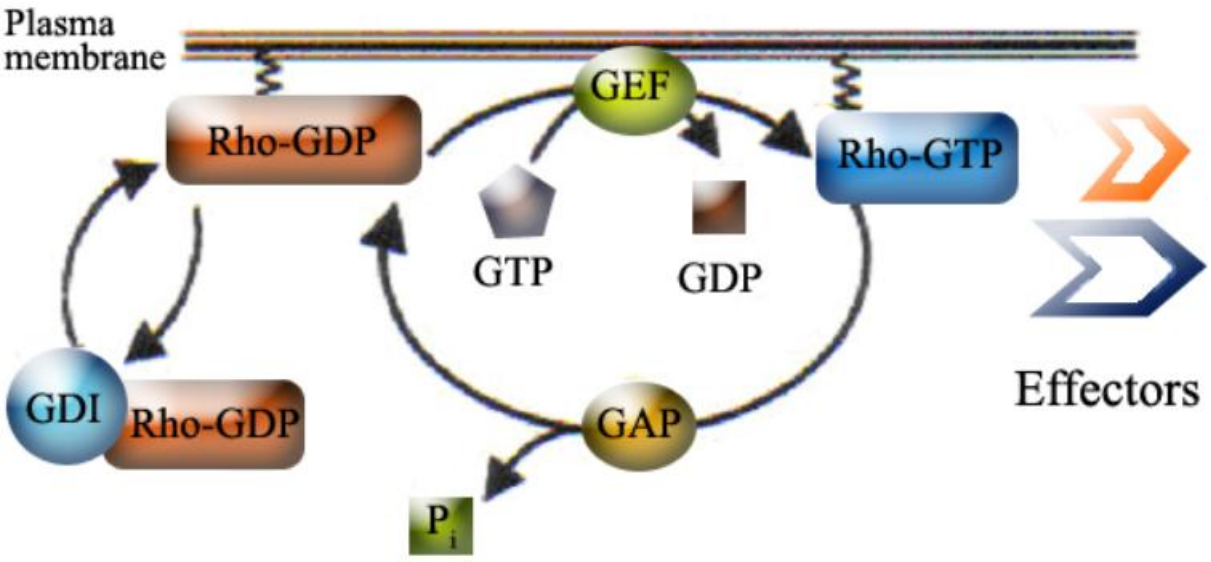


Figure 1.4: Regulation of small GTPases by GEFs, GAPs and GDIs. The activation cycle is shown for a member of the Rho family, although GAPs and GEFs act in a similar way for all other known small GTPases.

1.3.1.2 Structural basis for extracellular plexin-semaphorin signalling

Until recently, no structural information on the plexin ectodomain or its interaction with semaphorins was available and thus the determinants for extracellular specificity and the molecular mechanism controlling plexin-semaphorin signalling remained unclear. Three groups have now independently reported structures for several different plexin-semaphorin complexes that allow a detailed description of the interactions in the complex and suggest a signalling mechanism common to all plexin classes [35-37].

The structure of a truncated Plexin-A2 construct revealed that the seven-bladed β-propeller that forms the sema domain is more closely related to the Met receptor than to semaphorins. The reported crystal structures of PlexinA2-Sema6A, PlexinB1-Sema4D, PlexinC1-Sema7A and PlexinC1 in complex with the viral semaphorin mimic A39R all share the same overall architecture (Fig. 1.5). In the complex, semaphorin forms a homodimer, as observed in other uncomplexed semaphorin structures [67,68], and brings together two plexin monomers resulting in a symmetric 2:2 complex. The proteins interact in a head-to-head fashion through their sema domains with almost no interaction between the two plexin monomers. The C-termini of plexin and semaphorin point into opposite directions therefore allowing for this architecture to be suitable for signalling between apposing cell surfaces.

The same edge-on face of the β-propeller of the sema domain in plexins and semaphorins is involved in their interaction. Distinct insertions in particular between blades 1 and 2 and a long insert within blade 5, which has previously been described as the extrusion [68], are responsible for the molecular recognition of cognate receptor-ligand pairs.

The architecture of the complex with one semaphorin dimer bringing together two plexin monomers suggested a possible role of bivalency that would strengthen their interaction [69]. Indeed, extensive studies of the binding affinity using Surface Plasmon Resonance (SPR) revealed a strong bivalency effect that is dependent on the dimerization of semaphorin [35]. Dimeric semaphorin binds about 40-fold tighter to plexins than a monomerized mutant (Fig. 1.5). Interestingly, dimerization of semaphorin was also shown to be essential for plexin function since monomerized semaphorin fails to induce the typical plexin-induced collapse response in COS cells even at concentrations above its dissociation constant [35] (Fig. 1.5). In summary, these results suggest that dimerization of the N-terminus of the plexin ectodomain as a result of bivalent semaphorin binding is a prerequisite for plexin signalling. The dimerization could then seed further oligomerization either through plexin-plexin contacts on the outside or within the intracellular domain.

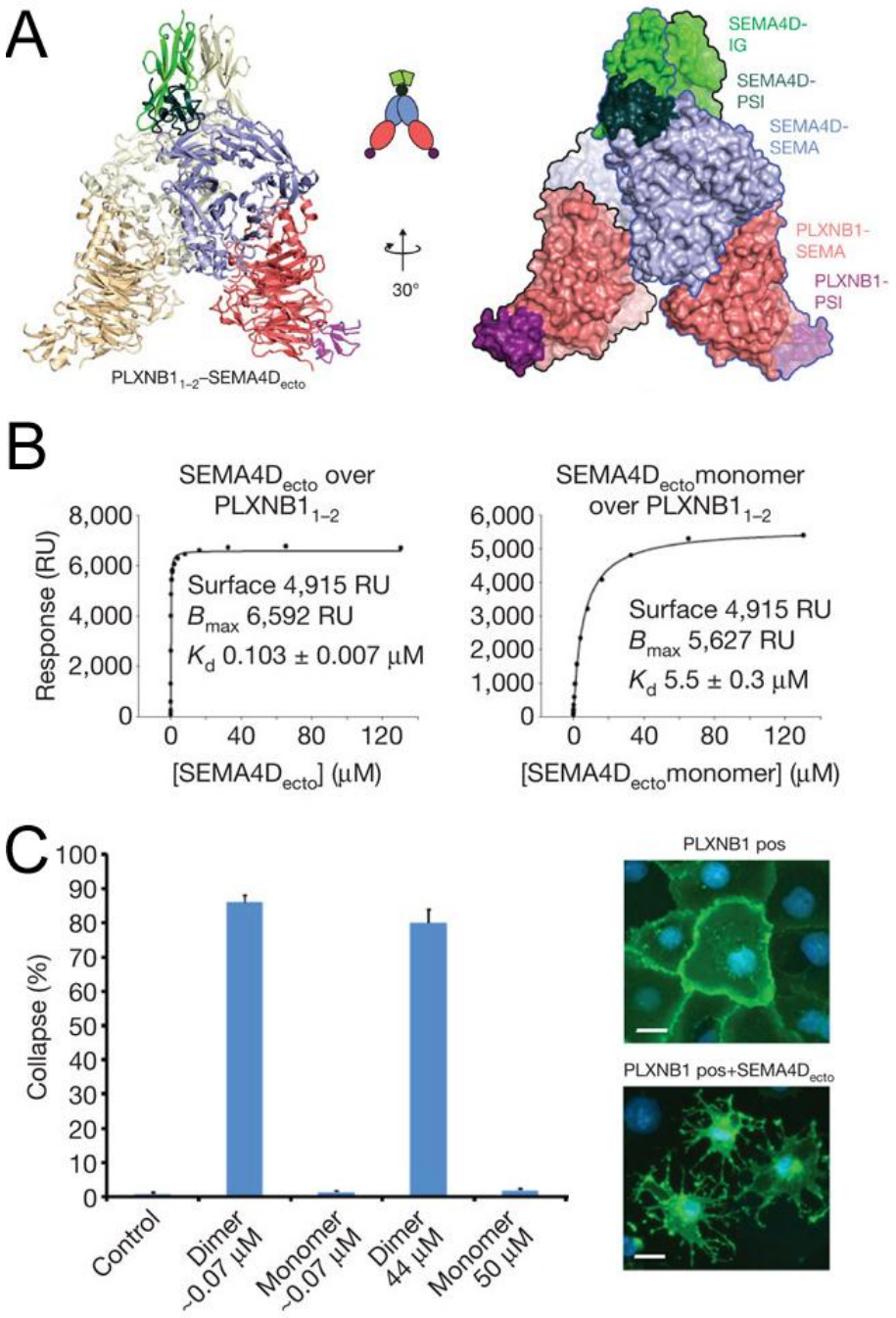


Figure 1.5: Structural and functional characterization of the semaphorin-plexin ectodomains. (A) Ribbon representation (left panel), cartoon drawing (middle panel) and surface representations with individual protein chains indicated by an outline (right panel) of the PlexinB1–Sema4D complex. (B) SPR equilibrium experiments of the interaction between PlexinB1 and wild-type and monomerized Sema4D, respectively. RU, response units. (C) COS-7 cell collapse assay showing representative images of non-collapsed cells (right, upper panel) and SEMA4D$_{ecto}$-induced collapsed cells (right, lower panel). Scale bar, 40 µm. (adapted from Janssen et al. [35])

1.3.2 Repulsive guidance molecules

Repulsive guidance molecules (Rgms) are recently discovered glycoproteins that contain a glycosylphosphatidylinositol (GPI) anchor that localizes them to the cell membrane [70]. The Rgm family consists of three members RgmA, RgmB and RgmC that share sequence homology of ~40–50 % and also have similar predicted protein domains and overall structure [71]. Each protein undergoes different posttranslational modification events including proteolytic cleavage which leads to distinct functions for each family member. In addition, the expression pattern of each of the three proteins is tissue-specific which further strengthens their distinct roles in signal transduction pathways ranging from repulsive axonal guidance (RgmA) to immune cell regulation (RgmB) and regulation of systemic iron metabolism (RgmC).

Rgms play important roles in the regulation of the Neogenin signalling pathway, which is crucially involved in axon guidance and cell survival [34,72]. Rgm family members can bind directly to Neogenin, however, the molecular mechanism of receptor activation and signal transduction remains unclear.

Similarly, all Rgm proteins are capable of binding selected bone morphogenetic proteins (BMPs). It is well-established that the role of RgmC in iron homeostasis is dependent on its interaction with BMPs, however it is only poorly understood what role RgmA and RgmB play in this important morphogenetic pathway. There is no structural information on any Rgm family member since none of the proteins show homology to any known structure.

1.3.2.1 RgmA

RgmA was the first family member to be identified [70]. It was cloned initially from mRNA isolated from chicken embryonic optic tectum. Subsequently, RgmA has been found in ten mammalian, eight non-mammalian and several invertebrate species [73]. Its mature form consists of two chains derived from a primary translation product of about 430 residues that form one integral structural unit due to disulfide bonds connecting the two separate chains. It also includes a C-terminal GPI anchor that localizes the protein to the cell membrane. The mechanism of intramolecular cleavage has not been established although it has been suggested that cleavage occurs during secretion at an acid-labile Asp-Pro bond [74]. RgmA is Cys-rich with 14 Cys residues whose disulfide pattern has not been established yet. The protein is glycosylated with three predicted N-linked glycosylation sites in mammals and two in chicken [70]. It is not known if other isoforms like single chain species or secreted forms of the protein exist.

RgmA was identified as a chemorepellent in temporal retinal axons of the developing chicken retina [70]. Interestingly, a knock-out of RgmA in mice did not have any effects on the retinal system but rather lead to defects in neuronal tube closure [75]. Thus the exact function of mammalian RgmA *in vivo* is not known. In addition to the role of RgmA in the central nervous system (CNS), it has also been identified as tumor suppressor in colorectal cancer [76]. Furthermore, recent studies showed that RgmA can modulate T cell responses and is involved in autoimmune encephalomyelitis, thus being a promising molecular target for the treatment of multiple sclerosis.

The receptor for RgmA in axon guidance is Neogenin. Neogenin is a type I transmembrane protein that also binds to the axon guidance molecule Netrin. Unlike Netrins, which bind to receptors related to Neogenin like Deleted in Colorectal Cancer (DCC) and have either attractive or repulsive effects depending on the receptor they bind to, RgmA exclusively binds to Neogenin and induces a chemorepulsive response. In addition, the RgmA-Neogenin interaction has been found to prevent neuronal apoptosis [34]. The molecular mechanism underlying these effects has not been established yet, however, small GTPases of the Rho family, in particular RhoA, and small GTPase regulators like RhoA kinase and p120 GAP (via focal adhesion kinase) have been suggested to be involved [77-79].

RgmA has also been found to bind specifically to the BMP ligands BMP-2 and BMP-4 but not BMP-7 or the transforming growth factor (TGF)-β 1 protein [80]. Binding results in enhancement of BMP signalling in a ligand-dependent manner [81].

1.3.2.2 RgmB

RgmB was discovered initially by a bioinformatic search for RgmA homologues [75]. It can be found as a single-copy gene in eight mammalian and seven non-mammalian vertebrates. The expression pattern of RgmB does not overlap with that of RgmA and it could be co-localized with DRG11 in dorsal root ganglia (DRG) and the spinal cord, leading to its additional name Dragon (DRG-ON) [82]. When compared with RgmA, RgmB shares sequence identity of ~ 50%. Like RgmA, RgmB is a GPI-anchored glycoprotein.

RgmB is suggested to play a role in cell adhesion due to its possible ability to form homophilic interactions [83]. It was also shown that RgmB, like RgmA, binds to selected BMP

ligands and acts as co-receptor enhancing BMP signalling [84]. In this context, RgmB was found to be involved in the regulation of immune responses by negatively regulating IL-6 expression in a BMP ligand dependent manner [85]. Furthermore, a strong interaction between Neogenin and RgmB has been shown and functional studies suggest that this interaction might be involved in neuronal migration in the dentate gyrus [86].

1.3.2.3 RgmC

RgmC was independently discovered in three different contexts. First, as a gene that resides in a locus that is involved in the human iron homeostasis disorder juvenile hemachromatosis, second due to homology to RgmA and RgmB and third as an novel protein expressed in skeletal muscle differentiation [87-89]. It can be found in nine mammalian and six non-mammalian vertebrates, however, unlike for RgmA and RgmB, RgmC cannot be found in avian species including chicken.

RgmC is a GPI-anchored glycoprotein that contains the conserved Asp-Pro autoproteolytic cleavage site. In addition, RgmC has a pro-protein convertase recognition site near the C-terminus which can be cleaved by furin [90]. Due to these several possibilities for posttranslational modifications, RgmC undergoes a series of processing steps leading to a variety of different, mature proteins [91]. RgmC has been found as membrane anchored single-chain or two-chain protein, suggesting that the Asp-Pro cleavage site is not fully cleaved during secretion. In addition both species can be truncated at their C-terminus by furin, thus leading to soluble, secreted forms of RgmC.

The biological function of RgmC has been studied extensively in the context of the iron disorder juvenile hemachromatosis since mutations in the human RgmC gene could be linked directly to the disease [88]. Juvenile hemachromatosis is a severe and rapidly progressing hereditary disorder of iron overload. One of its phenotypes, and indeed RgmC deficient mice, is a reduced level of the secreted hepatic peptide hepcidin. Hepcidin regulates the uptake and release of iron from the duodenum and macrophages, respectively. Expression and secretion of hepcidin are regulated by BMP and RgmC has been suggested to promote hepcidin expression through interaction with BMP [92]. Recently it has been proposed that Neogenin is needed as a co-receptor for this interaction [93]. Potential roles for BMP-independent signalling of RgmC-Neogenin complexes have not been established yet.

1.3.2.4 Neogenin signalling

Neogenin is a type-I transmembrane protein belonging to the immunoglobulin (Ig) superfamily of receptors. It is homologous to DCC, a netrin receptor that plays a significant role in cell migration and axon guidance. The two receptors share a sequence identity of ~50 % and display the same secondary structure. The ectodomain consists of four Ig domains followed by six fibronectin type III (FNIII) domains. The cytoplasmic tail comprises three conserved domains referred to as P1, P2 and P3.

Both Neogenin and DCC bind to netrins. Netrins are highly conserved, secreted proteins that share structural homology with laminins [94]. They can act both as repulsive and attractive guidance cues via a unique signalling mechanism. Attraction is mediated by binding of netrin

to DCC, causing clustering of multiple DCC receptors through their intracellular P3 domain. In repulsion, netrin simultaneously binds to DCC and a co-receptor of the UNC5-family therefore triggering interaction of their intracellular domains leading to long-range repulsion. It has been suggested that netrins are responsible for Neogenin-mediated chemoattraction in neuronal development and angiogenesis [95,96] and can also provide an adhesive rather than guidance function [97].

Unlike netrins, Rgms bind specifically to Neogenin and not DCC [72]. The binding of Rgm to Neogenin is unaffected by netrin since both ligands have independent binding sites on the receptor. Rgm binds to the fifth and sixth FNIII domain whereas netrins bind to the fourth and fifth FNIII domain [74,98].

Neogenin also plays a role in the formation of muscle junctions. Here, a direct interaction with CAM-related/down-regulated by oncogenes (CDO) is required for netrin-dependent myotube formation [99]. Neogenin and CDO interact with their ectodomains and form a heterodimer which then induces higher order oligomers with M- and N-cadherins.

On the intracellular side little is known about the signalling cascades triggered by neogenin activation. Initial experiments suggest that neogenin and DCC employ similar pathways in response to netrin binding. Both Neogenin and DCC bind to focal adhesion kinase (FAK). FAK is bound to the C-terminal half of the P3 domain of the receptor and upon netrin binding the receptors become phosphorylated on a tyrosine. This phosphorylation event recruits other src kinases which continue to phosphorylate FAK and themselves [100]. Interestingly, Rgm-mediated stimulation of Neogenin results in dephosphorylation of FAK. This allows dissociation of the kinase from p120GAP which is followed by an increase in interaction

between p120GAP and Ras-GTP. Thus RGMa binding to neogenin leads to the inactivation of Ras, which is required for the RGMa-mediated repulsive function in cortical neurons [78].

In addition to regulation of Ras, Neogenin can regulate small GTPases of the Rho-family thereby triggering dynamic rearrangements of the actin cytoskeleton and thus axon repulsion or attraction. The chemorepulsive Rgm-Neogenin interaction is mediated by RhoA activation of RhoA kinase (ROCK) [101]. Activated ROCK can then phosphorylate myosin, leading to F-actin depolymerisation and inhibition of neurite outgrowth [102]. The chemoattractive netrin-DCC interaction has been shown to result in Rac1 and Cdc42 activation [103]. Since netrin promotes the same chemoattractive response when bound to Neogenin it is likely that this is achieved by activation of the same Rho-GTPases. Interestingly FAK and src-kinases can directly activate Rac1 and Cdc42, thus the phosphorylation of these kinases in a Neogenin-dependent way might provide a molecular mechanism in which growth cone extension and attraction are triggered by netrin-Neogenin interactions via Rac1 and Cdc42 [104].

1.3.2.5 BMP signalling

Rgms also play an important role in the bone morhogenetic protein (BMP) signalling, a fundamental pathway in embryonic development. BMPs belong to the transforming growth factor-β (TGF-β) family and are crucial morphogens involved in tooth, kidney, skin, hair, muscle, haematopoietic and neuronal development [105]. Dysfunctions in BMP signalling lead to developmental defects, disease and cancer [106]. They were first identified as inducing

factor for bone and cartilage formation and today are used as therapeutics for complicated fractures.

Signal transduction by BMPs is initiated by binding to two different receptor types, referred to as type I and type II that both represent transmembrane serine-threonine kinases. Binding of the BMP dimer leads to receptor oligomerization with each BMP monomer being able to bind a type I and type II receptor simultaneously (Fig. 1.6). Upon receptor oligomerization the constitutively active type II receptor phosphorylates the type I receptor thereby activating the type I receptor kinase [107]. The type I receptor can then phosphorylate and activate SMAD proteins, which translocate into the nucleus where they, in concert with other proteins, act as transcription factors to regulate responsive genes [108]. It is important to note that the transmembrane receptor directly regulates the activation of the SMAD transcription factors therefore providing a pathway that leads directly from the cell surface to the nucleus. There are two different SMAD pathways that can be activated by members of the TGF-β superfamily. Which of the two pathways is activated depends on the capabilities of the individual type I receptor to phosphorylate either SMAD-2/3 or SMAD-1/5/8 [109,110]. The molecular determinants for the different interactions lie on a short, intracellular loop region of the type I receptor.

BMP signal transduction via type I and type II receptors is further regulated by extracellular, secreted antagonists (e.g. Gremlin), membrane-bound co-receptors and pseudoreceptors (e.g. Bambi) as well as intracellular inhibitor molecules (e.g. R-SMADs). Rgms are co-receptors for BMP-2, BMP-4 and BMP-6 and enhance BMP signalling. In addition it has been shown that the presence of Rgm can alter the receptor specificity of BMPs, facilitating the

binding of AcvRII instead of the canonical type II receptor BMPRII [80]. Furthermore, Neogenin has also been shown to be important for RgmC-mediated activation of BMP-signalling by localizing RgmC to the membrane [93] and for chondrogenesis both in vitro and in vivo [111].

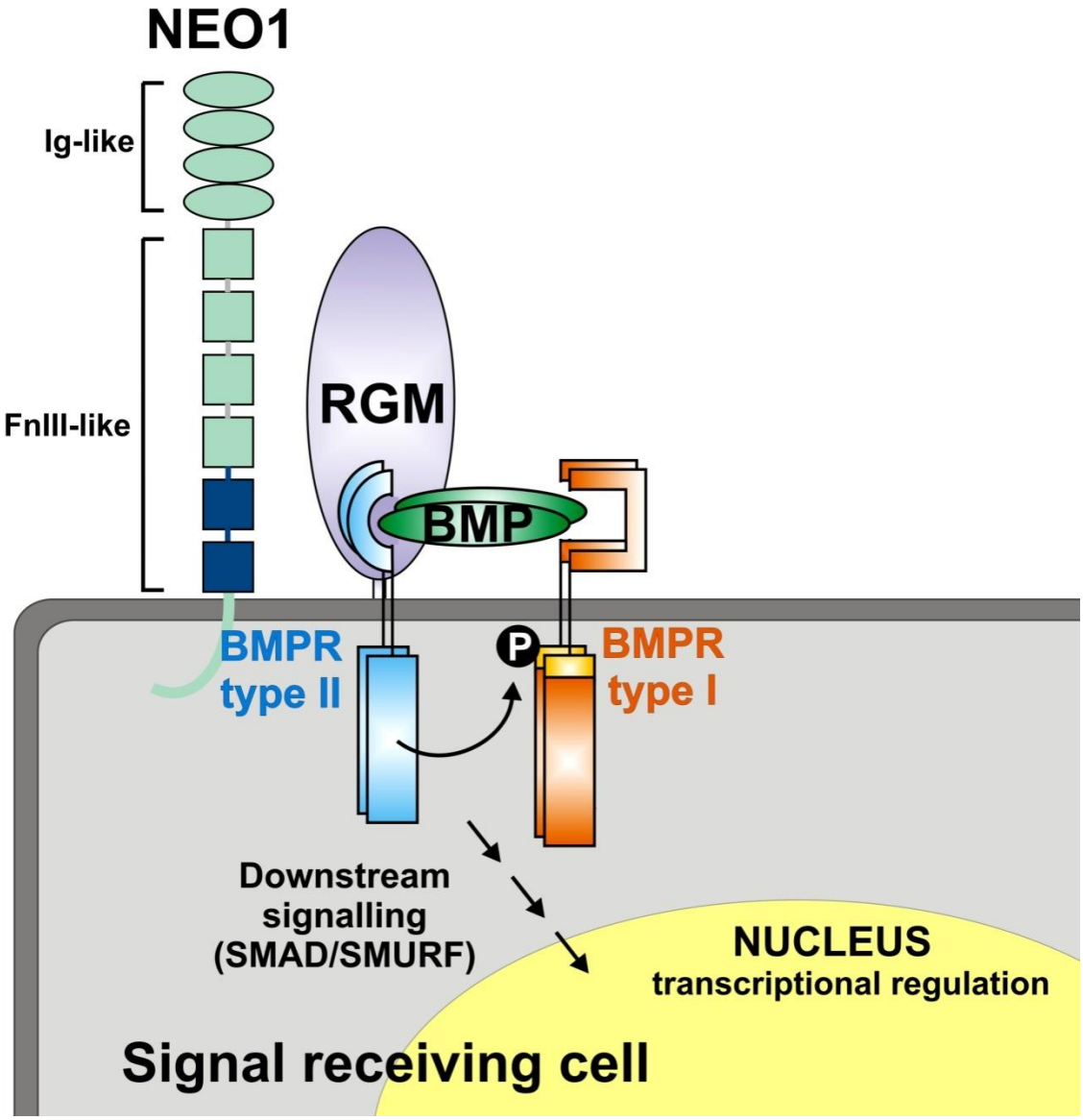


Figure 1.6: Schematic overview of BMP signalling. Dimeric BMP ligands oligomerize type I and type II receptors leading to transphosphorylation of the receptor and activation of the transcription regulator SMAD. Rgm is a co-receptor for BMP that enhances BMP signalling and Neogenin, the Rgm receptor, has been found to be involved in this function.

1.4 Aims

Despite extensive research, the signal transduction mechanisms underlying chemotaxis in bacteria and eukaryotes are still not fully understood. In this thesis two-component signalling in bacteria as well as the plexin-semaphorin and the Rgm-Neogenin pathway in eukaryotes were studied extensively. Structural and biophysical characterization combined with functional data using *in vivo* assays are employed to reveal the intricate mechanisms underlying molecular recognition and signal transduction mechanisms.

Chapter 2

Materials and Methods

Chapter 2 Materials and Methods

2.1 Molecular Biology

2.1.1 cDNA and expression vectors

$CheA_3P1$ (residues 1-135 of $CheA_3$, Gene Bank ID 3720125), $CheY_6$ (Gene Bank ID 3720126), $CheY_4$ (Gene Bank ID 3722004) and $CheY_3$ (Gene Bank ID CP000143) were cloned into the bacterial expression vector pQE80 (Qiagen), which includes an N-terminal His_6-tag. Human Rac1 (residues 1-176, GenBank ID: CAB53579) and human Rnd1 (residues 5-200, GenBank ID: BAB17851) were cloned into the bacterial expression vector pET22b, including a C-terminal His_6-tag. Human BMP-2 (residues 283-396, GenBank ID: NG023233) was cloned into pET22b without the C-terminal tag.

A series of constructs of the intracellular domain of human Plexin-B1 (GenBank ID: NP_001123554) lacking both C- and N-terminal regions as well as the RBD were designed and cloned into pBacPAK9 with a C-terminal His_6-Tag for purification.

For expression in mammalian cell lines the vectors pHLSec and pHLAvitag3 were used [112]. Human RgmA (GenBank ID AL136826), human RgmB (GenBank ID AK074887), human RgmC (GenBank ID AY372521), mouse Neogenin (GenBank ID Y09535) and their truncated or mutated constructs were cloned into these vectors introducing a Kozak sequence at the N-terminus, an N-terminal secretion signal and a C-terminal His_6-tag. Proteins used in Surface Plasmon Resonance, including all plexin mutants, were cloned into pHLAvitag3 introducing a

C-terminal biotinylation tag. For the intracellular plexin constructs the N-terminal secretion signal was left out. All constructs were cloned using primers purchased from MWG Operon.

2.1.2 Construct design and PCR

All expression constructs were designed with the aid of tertiary structure prediction using HHPred (http://toolkit.tuebingen.mpg.de/hhpred) and the disorder prediction server RONN [113]. Typically, PCR reactions were carried out in a total volume of 50 µl including: template DNA (approx. 2 ng/µl), 0.2 pmol/µl forward primer, 0.2 pmol/µl reverse primer, 1 x Pyrobest buffer II (containing 1 mM Mg^{2+}; Takara Co.), 0.4 mM dNTPs and 0.1 units Pyrobest DNA polymerase (Takara Co.). PCR cycles were performed using a Peltier Thermal Cycler-225 (MJ Research) as follows:

1. 96 °C for 120 s to melt the DNA template
2. 96 °C for 30 s to yield single stranded DNA
3. 55 °C for 30 s to allow the primers to anneal to the template DNA
4. 72 °C for 30-360 s elongation time depending on the length of the PCR product
5. Repeat steps 2-4 30 times
6. 72 °C for 10 min to ensure complete elongation

PCR products were then gel-purified using the QIAquick gel extraction kit (Qiagen) or in some cases purified using the QIAquick PCR purification kit (Qiagen).

2.1.3 Restriction digests, ligations and transformations

Purified PCR products and 2 µg quantities of the appropriate vector were digested for two hours with the appropriate restriction enzymes using the recommended buffer

concentrations and temperatures (New England Biolabs). The digested PCR products were purified, whereas the digested vectors were gel-purified. 20 µl ligation reactions were set up for 10 min at 25 °C with an insert to vector ratio of approximately 3:1 using the Quick Ligation Kit (New England Biolabs). 1 µl of the ligation mix was then used to transform 35 µl DH5α library efficient competent cells (Invitrogen); the cell-ligation mix was incubated for 30 min on ice and then heat shocked for 45 sec at 42 °C. After incubation of 2 min on ice, 100 µl SOC medium was added to the cells before incubation at 37 °C with 200 rpm shaking for 30 min. Finally cells were spread on LB-Agar plates containing 100 µg/ml carbenicillin and incubated overnight at 37 °C. Two colonies for each construct were picked and used to inoculate overnight cultures in order to prepare the plasmid DNA using the QIAprep spin miniprep kit (Qiagen). 1 µg purified plasmid DNA was digested using the appropriate restriction enzymes and analysed by agarose gel electrophoresis. Clones containing an insert of the correct size were then sequence verified (Geneservice, Oxford) and stored at -20 °C.

2.1.4 Site-directed mutagenesis

Oligonucleotide-based site-directed mutagenesis was performed on double-stranded DNA plasmid templates. For all constructs in the bacterial expression vectors the Quikchange Mutagenesis Kit (Stratagene) was used. Overlapping forward and reverse mutation primers were designed and a PCR reaction with the original plasmid as template was set-up. After the successful reaction the mixture was digested with DpnI, an endonuclease that specifically cuts methylated DNA. Since only the parental plasmid DNA but not the newly synthesized, PCR plasmid DNA is methylated this step leaves only plasmid DNA with the desired mutation. The mixture was then transformed and sequenced as described in 2.1.3.

Site-directed mutagenesis for constructs used in eukaryotic expression was carried out following a two-step, overlap-extension PCR. In the first step, two PCR products were synthesised: one consisting of sequences spanning from the mutation primer to the vector-derived 5' terminus primer and one consisting of sequences from the mutation primer to the vector-derived 3' terminus primer. Following the first PCR, the products were separated and purified by gel electrophoresis. The reaction mixtures for the second PCR were composed of the combined PCR products from the first PCR step (~200 ng), terminal vector-derived primers and standard Pyrobest PCR reagents. This reaction created PCR inserts with the desired mutation that could then be cloned into the appropriate vector.

2.2 Protein expression and purification

2.2.1 Protein expression and purification in *E. coli*

The expression plasmid containing the desired protein was transformed into M15 (for pQE constructs, Qiagen) or BL21 Star cells (for pET22b constructs, Invitrogen). Cultures were grown in Terrific Broth at 37°C until an OD_{600} of 0.8 was reached. Induction was carried out with 250 µM IPTG at 20°C over night. Cells were harvested by centrifugation (15 min, 7,000 g, 4 °C), resuspended in binding buffer (2 x phosphate buffered saline, pH 8, 500 mM NaCl, 0.5 mM β-mercaptoethanol) containing Complete protease inhibitor (Roche) and lysed by sonication (20 min, 9 s pulse on, 9 s pulse off, 60 % amplitude). The lysate was cleared by ultracentrifugation (60 min, 46,000 g, 4 °C) and the supernatant was supplemented with 2 mM $MgCl_2$ and incubated with 5 ml Ni-chelated sepharose resin (GE Healthcare) and 1000 Units DNase for 30 min at 16 °C. The beads were collected on a gravity flow column and washed with 2 x 50 ml binding buffer and 2 x 50 ml wash buffer (binding buffer + 5 mM

imidazole). The protein was eluted with 3 x 8 ml elution buffer (binding buffer + 300 mM imidazole). The eluted fractions were then concentrated and applied to a gelfiltration column equilibrated in 10 mM Hepes, pH 7.5, 150 mM NaCl, 2 mM TCEP.

For $CheY_3$ and $CheY_4$ the eluted fractions were pooled and dialysed against 2 l dialysis buffer (20 mM Sodium-Acetate, pH 4.5, 2 mM $MnCl_2$, 2 mM β-mercaptoethanol) for 1 hr at 4 °C. Precipitated protein was removed by centrifugation and the supernatant was concentrated to 1 ml and applied to a S75 16/60 gelfiltration column in dialysis buffer including 2 mM TCEP instead of β-mercaptoethanol.

BMP-2 was expressed as inclusion bodies and purified as follows adapting a protocol from von Einem *et al.* [114]. After cell lysis, the inclusion body pellets were washed with 60 mM EDTA, 6 % (v/v) Triton X-100, 1.5 M NaCl, pH 7, four times. The inclusion bodies were then solubilised in 6 M guanidinium hydrochloride (GdmHCl), 0.1 M Tris-HCl, pH 8.5, 1 mM EDTA and 100 mM DTT. The pH was then dropped to 3-4 to inhibit disulfide bond formation. Residual insoluble material was removed by centrifugation (10 min, 10,000 g, 4 °C). DTT was removed by dialysis against 10- to 20-fold volume of 5 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 hrs at 7 °C at a concentration of 200 µg/ml in 100 mM Tris-HCl, 5 mM EDTA, 1 M L-arginine, pH 8.3, 100 µM oxidized glutathione (GSSG), 100 µM reduced glutathione (GSH). The protein concentration was then increased by concentrating the sample to 1 mg/ml using Amicon pressure filtration (Millipore). The reaction was then split into two halves of which one was oxidized with 25 mM GSSG for 3 hrs at 7 °C. After this incubation, excess GSSG was removed by dialysis and the two reaction

halves were combined. The sample was then concentrated and applied to a heparin column followed by gelfiltration in 4 M Urea, 100 mM Tris-HCl, pH 6.0, 5 mM EDTA.

In all cases gelfiltration fractions were analysed by SDS-PAGE and those with a purity >95% were pooled and concentrated. The proteins were subsequently set-up for crystallization and remaining protein was aliquoted, flash-frozen in liquid N_2 and stored at -80°C.

For seleno-methionine labelling of proteins, cells were grown in minimal media (SeMet base, AthenaES) including 50 ml nutrient mix (AthenaES), 100 mg of each Lys, Thr, Phe, 50 mg of each Leu, Ile and Val and 50 mg L-seleno-methionine per litre. The addition of amino acids poisons the Met-synthesis so that only the seleno-labelled Met is incorporated into the proteins.

2.2.2 Protein expression and purification from insect cells

Constructs of the intracellular region of Plexin-B1 were cloned into pBac PAK9 (Clontech) including an N-terminal His_6-tag. Expression in insect cells was carried out by Weixian Lu (Strubi). A volume of 2 ml Sf9 cells (1 x 10^6 cells/ml) were transfected with the sequence verified vector. After five days the supernatant was collected and three rounds of virus amplification, each with a 1 : 100 dilution was carried out. For protein expression Sf9 cells with a density of 1.4 x 10^6 cells/ml are inoculated with the virus from the third round of amplification (1 : 10). After 3 days the cells were harvested by centrifugation. Protein purification followed the protocol established for bacterial proteins described in 2.2.1.

2.2.3 Protein expression and purification in mammalian cells

The HEK-293T cell line was mainly used for expression of the secreted Rgms, the different Neogenin constructs and of the plexin constructs used in SPR experiments. The mammalian cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM, Sigma), supplemented with non-essential amino-acids (Gibco), L-glutamine and 10 % fetal bovine serum (FBS, Sigma) at 37 °C, 5 % CO_2. All constructs were initially screened in small scale expression trials. Approximately 3 x 10^5 cells were seeded in 6-well dishes and grown to approximately 95 % confluency in DMEM containing 10 % FBS. Immediately before transfection the media was then exchanged into 1.5 ml DMEM, 2 % FBS, per well. Two 250 µl reaction mixtures, one containing 4 µg of plasmid DNA in serum-free DMEM, the other containing 8 µl Lipofectamin 2000 (Invitrogen) were mixed and incubated for 20 min at 25 °C. After this incubation the transfection mix was added to the cells. Cell media and cells were collected three days post-transfection and prepared for analysis by Western Blot to check for soluble protein expression.

The media was centrifuged to remove cell debris and the cells were resuspended in 1 ml PBS. Then, 10 µl of each sample were mixed with loading buffer and boiled for 5 min. These samples were then analysed on a sodium dodecyl sulphate polyacrylamide gel separated according to their size by electrophoresis (SDS-PAGE). For Western Blot analysis proteins were transferred onto a nitrocellulose membrane at a constant voltage of 30 V for 70 min using 10 % Methanol in NuPAGE Buffer and an XCell SureLock Mini-Cell (both Invitrogen). The membrane was blocked in 5 % milk in PBS for 15 min before probing with a monoclonal anti-pentaHis antibody produced in mouse (1:1000 dilution, Qiagen) for 1 h at 25 °C. After a

short wash step with PBS, 0.1 % Tween 20, the membrane was then probed with a goat anti-mouse antibody coupled to horseradish peroxidise (1:2000, Sigma) for 1 h at 25 °C. After multiple, extensive wash steps using PBS, 0.1 % Tween 20, the signal was detected using the enhanced chemoluminescence kit and exposure to Hyperfilm ECL (both GE Helathcare).

Large scale production of secreted protein constructs was carried out using HEK-293T cells, grown to about 90 % confluency in expanded surface roller bottles (2125 cm^2, Greiner Bio-One). These cells were transfected using a transfection mixture of 2 mg chloroform-extracted DNA (prepared using Endotoxin-Free Plasmid Mega or Giga kits, Qiagen) and 4 mg polyethylenimine (PEI) per litre of cell media. The roller bottles were incubated at 37 °C, 5 % CO_2 and constant, gentle rotation and cells or media were harvested 3-7 days after transfection. Alternatively, large scale expression was carried out by Benjamin Bishop using a custom-built TAP CompacT SelecT tissue culture robot, using HYPERflasks (1720 cm^2, Corning) instead of roller bottles.

Since all of the secreted proteins studied in this thesis were glycosylated, 5 µM Kifunensine (Toronto Research Chemicals) was added together with the transfection mixture. This specific inhibitor of α-mannosidase, stalls the N-linked glycosylation pathway producing a highly homogeneous population of $Man_9GlcNAc_2$ sugars [115]. These high mannose glycans can then be further processed by Endoglycosidase F1 leaving only a single GlcNAc residue.

Protein purification began by harvesting the cell media, followed by concentration and buffer exchange against PBS using a diafiltration device with a 5 kDa molecular weight cut off (GE Healthcare). The media was then incubated with 5 ml Talon resin (Clontech) for 1 h

at 16 °C. The remaining purification steps were identical to the ones for proteins expressed in *E. coli*.

2.3 Protein crystallization

Nanolitre crystallization trials using a Cartesian Technologies robot (100 nl protein solution plus 100 nl reservoir solution) were setup in 96-well Greiner plates, placed in a TAP (The Automation Partnership) Homebase storage vault maintained at 295 K and imaged them via a Veeco visualization system [116,117]. A pre-crystallization test (Hampton Research) was performed initially for each unique protein sample in order to determine an appropriate protein concentration to begin crystallization screening experiments. Crystal optimization was carried out in the same format as screening. Routinely, three different optimization strategies were followed. The first included addition of additives using the Hampton additive screen to the original crystal condition. The second employed a dilution series of the original condition and the last one, 3-row optimization, included a dilution as well as a change in pH of the original condition. To increase the size of crystals two different seeding strategies were employed, streak-seeding and micro-seeding. For the former protein crystals were touched with a cat's whisker which subsequently was passed through pre-equilibrated crystallization drops containing the original crystallization condition and varying concentrations of protein. For micro-seeding experiments, existing crystals were collected in approximately 50 µl of the reservoir solution and then processed using a seed bead (Hampton Research). This micro-seed stock was then used to inoculate a 96-well plate containing 100 nl of a dilution of the original condition and 100 nl protein.

2.4 Basic principles of crystallography

2.4.1 Description of crystals and Miller indices

A crystal can be depicted as a periodic, three-dimensional pattern of atoms or molecules. The description of the geometry of a crystal is not affected by the nature of the molecules or atoms it consists of. Thus, every molecule or atom can be treated as a single mathematical point. Taking this into consideration, a crystal consists of a three-dimensional ordered array of points, which referred to as lattice. The lattice exhibits translational periodicity in all three spatial directions. The smallest unit with which this periodicity can be achieved is called the unit cell. It contains the full symmetry of the crystal.

The unit cell is defined by three vectors *a*, *b* and *c* for each spatial direction and the three angles α, β and γ between them. It can be subdivided in an infinite number of lattice planes, which are defined by one point on each vector (Fig. 2.1). A single lattice plane can be translated along the lattice which leads to a set of parallel lattice planes that all have the same distance to each other. A set of lattice planes subdivides the vectors *a*, *b* and *c* into integer parts h, k and l which are called Miller indices. Every set of lattice planes is unambiguously defined by these indices.

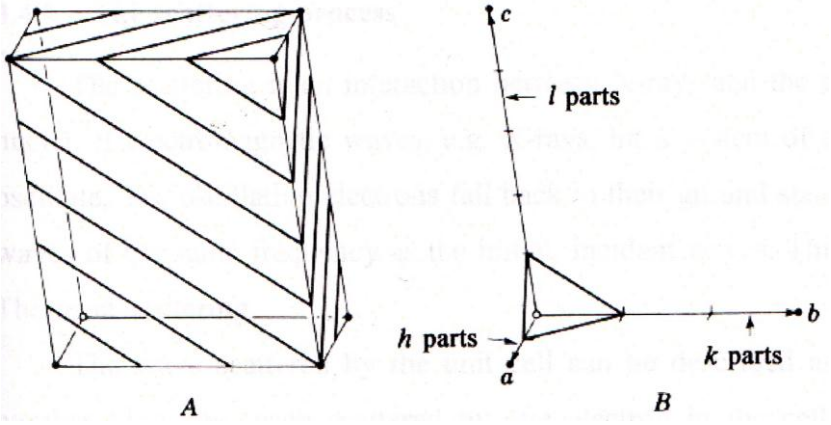


Figure 2.1: Example of a set of lattice planes. (A) The (235) set of lattice planes; (B) Definition of Miller indices for the same set as in (A).

2.4.2 Crystal symmetry

In addition to the translational symmetry among the unit cells in one crystal, additional symmetry elements can be found within one unit cell. These include inversion centers, rotation axes, mirror planes, glide planes, screw axes and rotation inversion axes. The combination of these elements leads to 230 space groups, which can be found in the *International Tables for Crystallography*. Due to the asymmetric nature of biological macromolecules, only 65 space groups are available for these molecules. The most common space groups for proteins are monoclinic $P2_1$ and orthorhombic $P2_12_12_1$. These space groups allow the entropically most favourable crystal packing.

2.4.3 X-rays

X-rays are electromagnetic radiation with wavelengths (λ) of 10^{-11} – 10^{-9} m (0.1 – 10 Å). Depending on their origin, they are called characteristic X-rays or Bremsstrahlung. Bremsstrahlung shows a continuous spectrum and does not have a distinct wavelength. In contrast, characteristic X-rays have a distinct wavelength. They are emitted after a series of

events. Firstly, an electron from one of the inner shells of an atom is displaced, e.g. by collision with a free, high energy electron. An electron from an outer shell replaces its position and during this process X-rays with a characteristic wavelength are emitted. This wavelength depends on the inner shell from which the initial electron was displaced from and the outer shell from which it was replaced. It is characteristic for every element. In crystallography, CuKα radiation is mostly used because of its suitable wavelength of 1.5418 Å. However, at synchrotrons, the wavelength used is more variable and usually is tuned to take advantage of the heavy atom (e.g. selenium) incorporated into the protein crystal for experimental phase determination.

2.4.4 The scattering process

The scattering process is an interaction between X-rays and the electrons surrounding the atomic nuclei. If electromagnetic waves, e.g. X-rays, hit a system of electrons, these electrons start to oscillate. The oscillating electrons fall back to their ground state by emission of electromagnetic waves of the same frequency as the initial, incident wave. The elastic scattering is also called Thomson scattering.

The wave scattered by the unit cell of a crystal can be described as a sum of the enormous number of waves, each scattered by one electron in the cell. The different positions of the electrons in the unit cell cause the scattered waves to cover different path lengths which means they are out of phase. For example, the phase shift caused by two electrons that are separated by the vector *r* is calculated as follows:

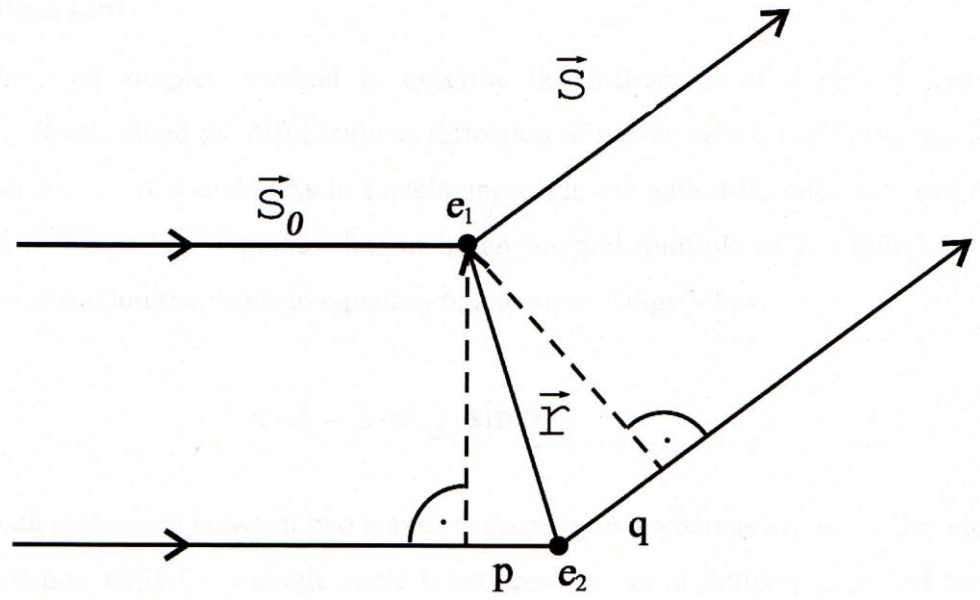


Figure 2.2: Scattering in a system with two electrons, e_1 and e_2. The path difference between the scattered waves is p + q.

An X-ray wave indicated by the vector s_0 with length λ hits the electron e_1 and is diffracted in the direction of the vector s with the same wavelength λ. Another wave is analogously diffracted by electron e_2. The path difference between the diffracted waves can be seen as a projection of r onto s_0 and s, respectively, multiplied by λ, which leads to:

$$p = \lambda \cdot \vec{r} \cdot \vec{s}_0 \qquad \text{(Eq. 1)}$$

$$q = -\lambda \cdot \vec{r} \cdot \vec{s} \qquad \text{(Eq. 2)}$$

$$p + q = \lambda \cdot \vec{r} \cdot (\vec{s}_0 - \vec{s}) \qquad \text{(Eq. 3)}$$

With respect to the wave scattered by e_1, the phase of the wave scattered by e_2 is:

$$-\frac{2\pi \cdot \lambda \cdot \vec{r} \cdot (\vec{s}_0 - \vec{s})}{\lambda} = 2\pi \cdot \vec{r} \cdot \vec{S} \qquad \text{(Eq. 4)}$$

Only if the latter phase is an integer multiple of the wavelength λ, the two waves amplify each other and a diffraction maximum can be observed. This is also called Laue condition.

2.4.5 Bragg's law

Another, yet simpler, method to describe the diffraction of a crystal was found by W.L. Bragg. He described the diffraction as reflection of waves on a set of lattice planes, defined by the Miller indices h, k and l. As in Laue's approach, the path difference between two waves, reflected on different lattice planes, has to be an integral multiple of λ. Figure 2.3 shows the geometric construction that leads to equation 5 known as Bragg's law.

$$n \cdot \lambda = 2 \cdot d_{hkl} \cdot \sin \theta_{hkl} \qquad \text{(Eq. 5)}$$

The path difference between two waves is given by $n\lambda$, whereas d_{hkl} is the distance between two lattice planes. Only for a single angle Θ between the set of lattice planes and the incoming wave can this condition be fulfilled and a diffraction spot (maximum) be observed. The diffraction lattice is only dependent on the space group and the unit cell dimensions and not on the nature of the molecules within the unit cell whereas the intensity of the diffraction maxima are related to the contributions of all of the atoms (and solvent) in the crystal lattice.

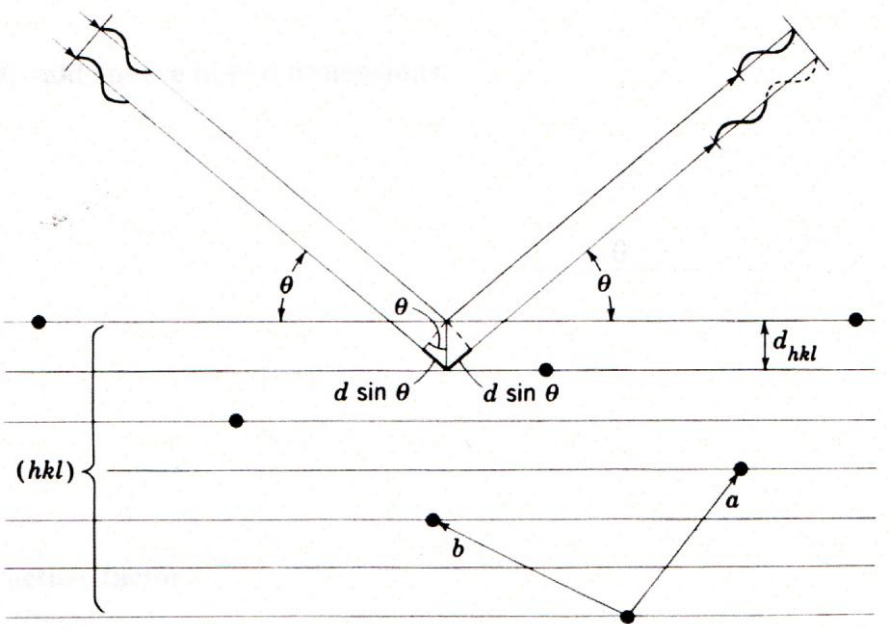


Figure 2.3: Geometric construction of Bragg's law.

2.4.6 The reciprocal space

Bragg's law clarifies the reciprocal relation between the scattering vector and the distance between the lattice planes, d_{hkl}. This relation can be described best with a reciprocal lattice. Every set of lattice planes, defined by the Miller indices h, k and l in the real space lattice, is transformed into a single point P_{hkl} in the reciprocal lattice. If the distance between P_{hkl} and the origin O* is equal to the reciprocal lattice distance d^*_{hkl}, Bragg's law is fulfilled and a reflection can be observed. This can be described geometrically by the construction of an Ewald's sphere (Fig. 2.4). If a reciprocal lattice point subtends the Ewald's sphere, Bragg's law is fulfilled. Every reciprocal lattice distance d^*_{hkl} (= 1 / d_{hkl}) can be described by a linear combination of the reciprocal vectors *a**, *b** and *c** and their integer multiples h, k and l. Thus, every reciprocal lattice point P_{hkl} can be described by Miller indices.

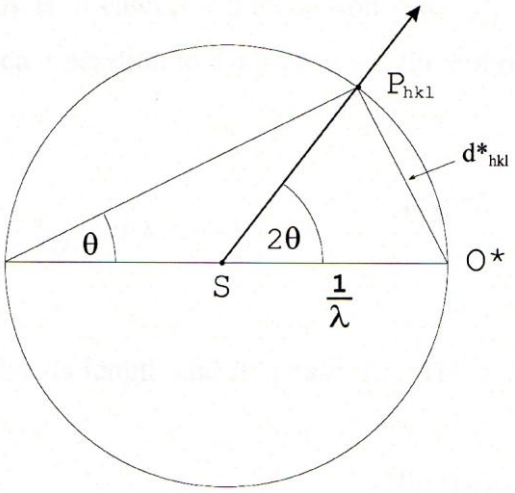


Figure 2.4: Ewald's sphere in two dimensions.

2.4.7 The structure factor

Every wave can be described as a vector in the complex plane. The amplitude of the wave is depicted by the length of the vector, whereas the phase of the wave is displayed by the orientation of the vector in the complex plane. If two or more waves are superimposed, these vectors are added and give rise to the structure factor. Each reflection is the result of the superposition of the diffracted waves, originating from every atom i, that fulfils Bragg's law. The structure factor is the sum of all these waves and can be calculated as follows:

$$\vec{F}_{hkl} = \sum_{i=1}^{n} f_i \cdot e^{2\pi i(hx_i + ky_i + lz_i)} \qquad \text{(Eq. 6)}$$

The atomic scattering factor f_i describes the ability of an atom to diffract electrons. It is dependent on the number of electron within that atom and is a function of 2(sinΘ)/λ.

2.4.8 Calculation of the electron density

Rather than taking the sum over every atom, one can also integrate over all electrons in the unit cell. The atomic scattering factor can then be expressed as electron density:

$$\vec{F}_{hkl} = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(xyz) \cdot e^{2\pi i(hx\mathbf{i}+ky\mathbf{i}+lz\mathbf{i})} dxdydz \qquad \text{(Eq. 7)}$$

However, the aim of protein X-ray crystallography is to calculate the electron density ρ at every position x, y and z in the unit cell. The mathematical operation to achieve this is the Fourier transformation of F_{hkl}, which leads to:

$$\rho(x,y,z) = \sum_h \sum_k \sum_l \vec{F}_{hkl} \cdot e^{-2\pi i(hx+ky+lz)} \qquad \text{(Eq. 8)}$$

Like every vector, the structure factor is defined by its length and its phase α_{hkl}. Thus, the electron density can be calculated as follows:

$$\rho(x,y,z) = \sum_h \sum_k \sum_l \left|\vec{F}_{hkl}\right| \cdot e^{[-2\pi i(hx+ky+lz)+i\alpha_{hkl}]} \qquad \text{(Eq. 9)}$$

The diffraction pattern of a crystal only yields the intensities I_{hkl} of every reflection hkl, which are equal to $|F_{hkl}|^2$. The phase angles αhkl cannot be measured. To solve this so-called phase problem several methods have been developed of which the following are the most commonly used:

i) Multiple Isomorphous Replacement (MIR)

ii) Multiple or Single Anomalous Dispersion (MAD/SAD)

iii) Molecular Replacement (MR)

In this thesis both SAD and MR phasing methods were employed and are described in section 2.6.

2.4.9 Friedel's law

The fact that the structure factor is depicted as a complex number leads to another symmetry operation. Reflections on a set of lattice planes from the opposite direction lead to the same modulus of the structure factor, but its phase has the opposite algebraic sign. Mathematically this can be described as:

$$\left|F_{hkl}\right| = \left|F_{\bar{h}\bar{k}\bar{l}}\right| \qquad \text{(Eq. 10)}$$

$$\alpha_{hkl} = -\alpha_{\bar{h}\bar{k}\bar{l}} \qquad \text{(Eq. 11)}$$

Because of this relation, all reflections of a crystal without any additional symmetry can be collected by rotating the crystal by 180 °.

2.4.10 Anomalous scattering

During the regularly observed Thomson scattering, the electrons are oscillating in direct response to an incident X-ray wave, which in turn produces elastic scattering of the X-rays. When the energy of the X-rays approaches the energy of an electronic transition from an atomic bound orbital (absorption edge), a resonance condition is established that amplifies the oscillation and perturbs the scattering. In X-ray crystallography, this phenomenon can only be observed for heavy atoms, in which the electrons of the inner shell cannot oscillate freely anymore. As a result, the atomic scattering factor becomes complex:

$$f = f_0 + \Delta f' + i\Delta f'' \qquad \text{(Eq. 12)}$$

The atomic scattering factor *f* is separated into a dispersive part *Δf'* and the absorption part *Δf''*. The absorption part is phase shifted by π/2. As a consequence, Friedel's law is no longer valid. This can be used to derive the location of the anomalous scatterers and solve the phase problem.

2.4.11 The temperature factor

The thermal movement of the atoms around their ideal, mean position is an important aspect to consider when describing three dimensional structures. Due to this movement, the amplitudes of the scattered waves are not equal in their phases leading to a decrease in the intensity of the reflection. The aforementioned decrease in intensity is dependent on the wavelength and the scattering angle 2Θ. The temperature factor B, or B-value, is related to the mean squared deviation of the position of the atoms as follows:

$$B = 8\pi\{u^2\} \qquad \text{(Eq. 13)}$$

The temperature factor for the whole crystal can be derived by plotting I_{hkl} / Σ f_i^2 against $\sin^2\Theta/\lambda^2$ in the so-called Wilson plot. The slope of the linear fit of this plot yields the mean temperature factor of all atoms in the crystal.

2.5 Data collection and processing

Diffraction data was collected at either the Diamond Light Source, Oxfordshire, UK, or the European Synchrotron Radiation facility (ESRF), Grenoble, France. For crystal screening and data collection of crystals diffracting to very high resolution even at low X-ray doses, a

laboratory X-ray source, consisting of Rigaku microfocus X-ray generator together with a mar345 imaging plate detector was used. At the synchrotron the data collection strategy function in Mosflm [118] was used to determine the optimal orientation of the crystal to ensure maximum completeness with minimum exposure. Typically, an oscillation range of 180 ° was collected. For Se-Met labelled protein crystals, a fluorescent scan was carried out to establish the exact wavelength at the absorption edge of selenium. Highly redundant data were collected on the peak of the absorption edge to enable phasing using SAD strategies. If the crystal did not suffer radiation damage, further datasets at the inflection point and at a high energy remote wavelength were collected to allow MAD phasing.

Data were processed using the HKL2000 package [119]. The diffraction images were first indexed to assign each reflection a Miller index and to determine the point group and unit cell dimensions of the crystal. This was followed by integration, merging and scaling of the intensities. Data quality and high-resolution limits were assessed by completeness, redundancy, the signal to noise ratio and the R_{merge}. The latter is a measure of the internal consistency of the dataset and represents the deviation between equivalent reflection spots. Structure factor amplitudes for each dataset were calculated from the reflection intensities using TRUNCATE as implemented in the CCP4 software package [120].

2.6 Structure solution and refinement

2.6.1 Phasing using Se-SAD

The $CheA_3$ P1-$CheY_6$ complex structure described in chapter 3 was phased by the single anomalous dispersion method using seleno-methionine as anomalous scatterer. After data collection the Se-Met sites were identified using autoSHARP [121]. AutoSHARP uses the ShelX

program suite to locate the position of the anomalous scatterer. Several statistics indicated a correct solution within ShelX. Firstly, in a plot of CC_{all}/CC_{weak} versus the trial number, correct solutions should lie in a separate cluster usually with CC values of above 35 %. Secondly, there should be a sharp drop of occupancy after the last correctly identified Se-Met site. AutoSHARP then uses the identified substructure to obtain initial phases. Since data were only collected at a single wavelength, a single cycle of density modification in autoSHARP was used to determine which of the two possible enantiomorphs represents the correct solution. There should be a significant difference in the figure of merit values obtained for the two possible solutions. In addition, the electron density maps generated after density modification should show characteristic protein features like α-helices or a sharp protein-solvent boundary for the correct solution.

2.6.2 Phasing using molecular replacement

Molecular replacement was performed in Phaser [122]. Phaser first performs a rotational search of the appropriate copies of the trial model and then uses the top solutions of this search to carry out a translational search. The top scoring solutions from the latter search are then ranked according to their number of clashes within the crystal lattice and the solution with the fewest clashes is then further refined using iterative rounds of rigid-body refinement producing the final solution statistics, models and maps. Two statistics give an initial indication of the correctness of a molecular replacement solution, the Z-score and the log-likelihood gain (LLG). The LLG represents a measure of how well the molecular replacement solution fits the experimental data compared to a randomly generated arrangement of the same atoms. The likelihood describes the probability that the calculated

model represents the experimental data and the LLG is the difference between this likelihood and the likelihood of the random distribution. The Z-score is the number of standard deviations above the mean LLG score for a particular LLG value. A correct molecular replacement solution usually has Z-scores of > 7 and a steadily increasing LLG. However, the clearest indication for a correct molecular replacement solution is the electron density map. If the density map shows positive density for missing features, e.g. side chains or even whole domains, the molecular replacement solution is likely to be correct. In addition, rigid-body refinement statistics should generate an R_{free} of < 50 %.

2.6.3 Refinement

For the high-resolution structures, the initial solutions obtained from either molecular replacement or experimental phasing were extended and corrected using the automatic model building program ARP/wARP [123]. The improved output models were then manually adjusted using localised real-space and geometry restrained refinement in COOT [124]. Refinement was carried out in autoBuster [125] and Phenix [126] using appropriate geometry weights of the atomic positions and B-factors for the entire model against the experimentally determined structure factors. The low-resolution models were built by iterative rounds of model building in Buccaneer and density modification in Parrot, both part of the CCP4 programme package [120]. These models were then manually adjusted in COOT [124] and refined using autoBuster [125] and Phenix [126].

Several parameters were monitored during the refinement process; R_{work}, R_{free} and the rmsd values of the bond lengths and the bond angles. R_{work} is a linear R-value that represents the deviation of the model structure factors in comparison to the data structure factor

calculated from all reflections used in the refinement process. R_{free} is calculated in a similar manner but only uses a random set of 5 % of the reflections that are not used in the refinement process. Thus R_{free} is a useful tool to cross-validate the model against the experimental data. A significant deviation of more than 5 % between R_{work} and R_{free} usually indicates the introduction of model bias in the working model. During iterative model building and refinement, both values should decrease continuously indicating a better fit between model and experimental data. The bond length and bond angle rmsd values describe the geometry of the model when compared to reference values from high resolution small molecules (Cambridge Structural Database). For low-resolution structures these values have to be restrained with a higher weight since there are not enough observations to account for all the variables in the structure. In high-resolution structures these restraints can be relaxed.

The structural models were further validated by assessing their stereochemical properties using Ramachandran plots and the validation tool MolProbity [127]. A Ramachandran plot shows the two backbone angles, Φ and Ψ, for each protein residue. There is only a limited number of allowed combinations thus if a large number of residues lies in a disallowed region of the plot this indicates poor model geometry.

Non-crystallographic symmetry (NCS) was imposed for all structures that contained more than one molecule in the asymmetric unit. The use of NCS operators reduces the number of refinement parameters significantly and is particularly helpful when refining low resolution structures. NCS restraints were relaxed in regions showing significant deviations between the individual units, e.g. due to different lattice contacts in the crystal.

TLS (translation-libration-screw) refinement allows a partial description of molecular anisotropy using only 20 parameters per TLS group compared to six parameters per atom for full anisotropic B-factor refinement. TLS refinement was used for all structures except the high-resolution $CheY_3$ structures where individual anisotropic B-factor refinement was employed. TLS groups were determined using the TLS Motion Determination server [128] and imported in the refinement program.

2.7 Surface Plasmon Resonance

2.7.1 Sample preparation

Two strategies were employed to couple protein ligands on the chip surface. For binding studies with BMP-2 as ligand, the protein was coupled to the chip via non-specific amine coupling. For all other measurements, strepatavidin was coupled to the chip via amine coupling; then the biotinylated ligand of interest was bound to the immobilized streptavidin. Due to the high affinity of this interaction with a K_d of approximately 10^{-14} M, binding of the biotinylated protein ligand is essentially irreversible. This two-step immobilization strategy has two main advantages over non-specific amine coupling. Firstly, the ligand is presented in a discrete orientation that can be modified by changing the biotinylation tag from one terminus to the other, rather than in random orientations of which some might obscure the binding site for the analyte. Secondly, the ligand does not have to be prepared in high purities since the streptavidin-biotin coupling provides an additional affinity purification step that separates unwanted impurities. With amine coupling all proteins with surface accessible Lys residues are coupled to the chip.

Ligand constructs were cloned either into pHLAvitag3 for expression in mammalian cells or a C-terminal biotinylation tag was introduced into the bacterial expression vectors. Ligands expressed in *E. coli* could be used for immobilization on the chip immediately after cell lysis. *E. coli* has an endogenous Biotin transferase and biotinylates the protein of interest during overexpression. For ligands expressed in mammalian cells, the cell medium or in the case of the intracellular plexin constructs the supernatant of a cell lysate was buffer exchanged into 10 mM Tris-HCl, pH 8.0, and concentrated to approximately 750 µl. The biotinylation reaction was started by addition of Biomix A buffer, Biomix B buffer (Avidity) and BirA enzyme to give the following final concentrations: 15 mM ATP, 15 mM Magnesium acetate, 80 µM biotin, 80 mM bicine, pH 8.3, 30 µg/ml BirA. The reaction mixure was incubated for 30 min at 30 °C and then buffer exchanged into the SPR running buffer (10 mM Hepes, pH 7.5, 150 mM NaCl, 0.05 % (v/v) Tween 20) using a PD-10 column (GE Healthcare).

2.7.2 Experimental setup

SPR experiments were performed using a Biacore T100 machine (GE Healthcare) at 25 °C in SPR running buffer. Protein concentrations were determined from the absorbance at 280 nm using calculated molar extinction coefficients. Typically the ligands were bound to the surface at concentrations of 150-3,000 response units and at a flow rate of 5 µl/min. After each binding experiment the chip was regenerated by short bursts of 50 mM EDTA, pH 8 (for all bacterial ligands) or 2 M Magnesium sulphate (for all other experiments). All experiments were done in duplicates with independently purified proteins. The signal from experimental flow cells was corrected by subtraction of a blank and reference signal from a mock or irrelevant protein coupled flow cell. In all experiments analyzed, the experimental

trace returned to baseline after each injection and the data fitted to a simple 1:1 Langmuir model of binding. K_d values were obtained by nonlinear curve fitting of the Langmuir binding isotherm (bound = $C^*max/(K_d + C)$, where C is analyte concentration and max is the maximum analyte binding) evaluated using the Biacore Evaluation software (GE Healthcare).

2.8 Phosphotransfer assay

This assay was performed by Dr. Steven L. Porter. Assays were performed at 20 ^oC in TGMNKD buffer (50 mM Tris HCl, 10% (v/v) glycerol, 5 mM $MgCl_2$, 150 mM NaCl, 50 mM KCl, 1 mM DTT, pH 8.0). $CheA_3P1$ was phosphorylated using [γ-^{32}P] ATP and $CheA_4$, and purified as described in section 2.2.1. The final preparation of $CheA_3P1$-^{32}P was free of ATP and $CheA_4$. 1 μM $CheA_3P1$-^{32}P was mixed with 5 μM of the response regulator in a final reaction volume of 100 μl. Following the addition of response regulator, reaction aliquots of 10 μl were taken at the indicated timepoints and quenched immediately in 20 μl of 1.5 X SDS-PAGE loading dye (3.75 % (w/v) SDS, 45 mM EDTA, 18.75 mM Tris HCl, 18.75 % (v/v) glycerol, 1.5 % (v/v) β-mercaptoethanol, pH 6.8). Samples were heated to 65 °C for 30 seconds prior to SDS-PAGE (15% (w/v) polyacrylamide gels). To minimize hydrolysis of phosphoproteins, electrophoresis was performed at 4 °C, with a run time of 45 minutes. Gels were dried, exposed to phosphor screens (Kodak) and analyzed using an SF-phosphorimager with ImageQuant software (version 5.0, Molecular Dynamics), correcting for the background present in each lane.

2.9 Swarm Plate assay

As a first step to study the $CheY_3$ mutants, *R. sphaeroides* had to be conjugated with the appropriate plasmids. The protein constructs were cloned into the *R. sphaeroides* expression vector pIND4 and subsequently transformed into S17-1 cells. Single colonies of a *R. sphaeroides* $CheY_3$-$CheY_4$ deletion strain (JPA1025) were picked and grown in the light cabinet at 30 °C. The transformed S17-1 cells were grown at 37 °C over night. 200 µl of the S17-1 cells and 1 ml of the R. sphaeroides cells were spun down and gently washed with 1 ml LB. The cell pellets were then resuspended in 100 µl LB, mixed and transferred to a sterile filter on a dry LB agar plate. After incubation at 30 °C for 24 hrs the filter was harvested, resuspended vigorously in 500 µl LB and the solution was spread on an LB agar plate carrying the appropriate antibiotics. Conjugated *R. sphaeroides* formed colonies after 48 hrs. Cells carrying the appropriate pIND4 expression plasmids were grown overnight in succinate medium with antibiotics at 30 °C. Aliquots (5 ml) were used to inoculate single sites on LB swarm plates containing 0.25% agar, appropriate antibiotics and IPTG at 0, 1, 10, 100 or 1000 mM. The plates were incubated at 30 °C. Swarm diameters were measured after 48 hrs. Each experiment was repeated nine times. At each IPTG concentration, the growth rates of the strains containing the expression plasmids were compared with those of the strains containing only the control vector. The non-motile control strain was JPA1301, and the non-chemotactic control strain was JPA1215.

2.10 Thermofluor assay

In order to assess protein stability under different buffer conditions and in the presence of additives a Thermofluor assay was carried out. In this assay the protein is mixed with a

SybrOrange, a fluorescent dye that binds to the hydrophobic core of a protein upon unfolding. This binding event changes the fluorescent spectrum of the dye and thus creates a folding-specific probe to study protein stability. The reactions were set-up in a 96-well format with a reaction volume of 50 µl. 5 µg of the protein were mixed with SybrOrange (final concentration 2 x) and the buffer or additive. The plate was then incubated in a RealTime PCR machine. The fluorescence was measured constantly and the temperature was increased in steps of 1 °C from 4 °C to 95 °C over a time course of 90 min. The resulting melting curves were then analysed using the manufacturer's software.

2.11 COS-7-cell collapse assay

Cellular collapse assays were performed following a protocol published earlier [129]. Briefly, COS-7 cells were seeded on glass coverslips and transfected with human Plexin B1 carrying an N-terminal Flag-tag. Two days after transfection, cells were treated with medium containing secreted wild-type or mutant Sema4D and incubated for 30 min at 37 °C. Finally, the cells were fixed and stained with anti-Flag primary antibody (Sigma) and Alexa 488-labelled secondary antibody (Invitrogen). Cell nuclei were counterstained with DAPI (Invitrogen) and cells were visualized with a TE2000U fluorescence microscope (Nikon) equipped with an Orca CCD camera (Hamamatsu). Plexin B1-expressing cells were classified as collapsed or non-collapsed on the basis of reduced surface area. Each experiment was repeated twice and 2×200 cells were counted each time. Results are shown as mean with error bars representing standard error of the mean.

2.12 BMP response assay

For the BMP response assay the mouse myoblast cell line C2C12 was used. Cells were stably transfected with a Luciferase reporter gene under the control of a BMP-inducible SMAD transcription factor. Cells were cultivated to approximately 90 % confluency in DMEM with 10 % FBS and then serum starved by exchanging the medium to DMEM containing 0.1 % FBS. Different concentrations of recombinant, purified BMP-2 were then added to the cells and incubated for 24 hrs. Then, the cell medium was removed; the cells were washed twice with PBS and then lysed using the lysis buffer supplied in the Luciferase Assay System (Promega). Cell lysates were then mixed with the luciferase substrate (Promega) and the luminescence was measured in a luminometer (BD Biosciences).

2.13 Analytical ultracentrifugation

Sedimentatrion velocity experiments were carried out on an Optima XL-I analytical ultracentrifuge (Beckman Instruments). Samples at different concentrations in 10 mM HEPES, pH 7.5 and 150 mM NaCl, 2 mM β-mercaptoethanol were centrifuged in double sector 12-mm centerpieces in an An-60 Ti rotor (Beckman) at 50,000 rpm and 20 °C. Protein sedimentation was monitored by Rayleigh interference. Data were analysed using SEDFIT [130].

2.14 Figures

Sequence alignments were performed using MultAlin [131] and formatted with ESPript [132]. Crystal structure images were generated with PyMol and pictures of electron density maps were generated with PyMol [133] or Coot [124]. For illustration of SPR sensorgrams and binding

curves, raw data was transferred to either Microsoft Excel or SigmaPlot (Systat Software). Other illustrations were prepared using Adobe Photoshop (Adobe Systems) or Corel Draw (Corel Corporation).

Chapter 3

Rewiring chemotaxis from the structure of a histidine phosphotransfer domain in complex with its cognate response regulator

Chapter 3 Rewiring chemotaxis from the structure of a histidine phosphotransfer domain in complex with its cognate response regulator

3.1 Summary

Two-component signal transduction pathways consisting of histidine protein kinases (HPKs) and their response regulators (RRs) are widely used to control bacterial behaviour, in particular, responses to environmental challenges. Some bacteria have over 150 different two-component pathways and the specificity of the phosphotransfer reactions within these systems is tightly controlled to prevent unwanted crosstalk. One of the best understood two-component signalling pathways is the chemotaxis pathway. In this chapter the crystal structure of the histidine-containing phosphotransfer domain of the chemotaxis HPK, $CheA_3$, in complex with its cognate RR, $CheY_6$ is presented. The determinants for specificity in this protein-protein interaction are deduced from the structural data and further validated by binding studies using Surface Plasmon Resonance and a functional phosphotransfer assay. These insights then allowed targeted mutation of non-cognate RRs to allow them to bind to and be phosphorylated by the HPK. In summary, the results presented here provide valuable insight into how cells mediate specificity in one of the most abundant signalling pathways in biology, two-component signal transduction.

3.2 Introduction

In bacteria, environmental stimuli are sensed by transmembrane as well as intracellular proteins that usually have a HPK at their core. The external stimulus controls the rate at which the HPK autophosphorylates on a conserved histidine residue. To relay the signal, the phosphorylated HPK transfers the phosphoryl group to a conserved aspartate residue within the receiver (REC) domain of the cognate RR. Upon phosphotransfer the RR, often a transcription regulator, changes its conformation and then effects a response appropriate to the original stimulus. In some bacteria there are over 250 different two-component pairs, each regulating specific events. Thus, the specificity between these HPK-RR pairs has to be tightly controlled to prevent crosstalk between the different pathways. There are several mechanisms that ensure this specificity, however, the most important one is molecular recognition where an HPK shows strong kinetic preference for its cognate RR [134,135]. Understanding this preference on an atomic level will not only contribute to predicting interacting HPKs and RRs but will also allow for rewiring of novel two-component pairs useful for synthetic biology.

Many two-component systems utilize an additional element, the Hpt (histidine-containing phosphotransfer) domain in phosphotransfer; these include multistep phosphorelays and chemotaxis signaling pathways. In multistep relays, the Hpts can serve both as phosphoacceptor and phosphodonor for RRs. They provide the intermediate stages from the initial phosphotransfer by the autophosphorylated HPK to the final output RR. In chemotaxis signalling pathways, the Hpt domain, also called P1, is part of the chemotaxis protein CheA. It is phosphorylated by the kinase domain (P4) of the same protein. Once

phosphorylated the Hpt then serves as initial phosphodonor for the interacting chemotactic RRs [136]. Structurally, Hpts are different to HPKs since they probably do not form dimers. However, the core of the dimeric HPK and the monomeric Hpt, a four-helix bundle, is remarkably conserved across both families [137].

In this chapter the structure of the Hpt (P1) domain of $CheA_3$ from *R. sphaeroides* in complex with its cognate response regulator $CheY_6$ is presented. This structure presents the first structure of a Hpt domain of a CheA protein in complex with its response regulator.

Unlike the well-understood chemotaxis pathway in *E. coli*, many bacteria including *R. sphaeroides*, have multiple homologues of each of the chemosensory proteins [138,139]. *R. sphaeroides* has four CheA homologues and eight chemotaxis RR proteins (six CheYs and two CheBs). In accordance with the tightly controlled specificity each of the CheA proteins shows different specificities for the RRs [24,25,140,141]. Interestingly, $CheA_3$ and $CheA_4$ combine to form a cytoplasmic signalling cluster which is supposed to sense the metabolic state of the cell [138 142,143]. In addition to their unusual interaction, $CheA_3$ and $CheA_4$ also lack some of the domains found in the *E. coli* CheA [24 144]. Whereas $CheA_4$ consists of a P3 (dimerization) domain, a P4 (kinase) domain and a P5 (regulatory) domain, $CheA_3$ comprises of a P1 (Hpt) domain and a P5 (regulatory) domain separated by a 794 amino acid sequence that encodes a novel $CheY_6$-P phosphatase activity [23]. None of the two CheA proteins are able to autophosphorylate. Instead, $CheA_4$ phosphorylates the P1 domain of $CheA_3$ which then serves as specific phosphodonor for the cognate chemotaxis RRs, $CheY_1$, $CheY_6$ and $CheB_2$. All other chemotaxis RRs, $CheY_2$, $CheY_3$, $CheY_4$, $CheY_5$ and $CheB_1$ are not phosphorylated. This phosphotransfer specificity must be determined by the interactions between the RRs

and the P1 domain of $CheA_3$ since the isolated P1 domain of $CheA_3$ ($CheA_3P1$) shows the same phosphotransfer specificity as full-length protein [23].

The aim of this chapter will be to define the molecular determinants for specificity in the $CheA_3$-$CheY_6$ complex and once determined, to use this knowledge to design novel, synthetic two-component pairs.

3.3 Expression and purification of $CheY_6$ and $CheA_3P1$

Constructs containing the full length response regulator $CheY_6$ (residues 1-134) and the P1 domain of $CheA_3$ (residues 1-135) were cloned into the bacterial expression vector pQE60 (see section 2.1). Both proteins were expressed in M15 cells following the protocol described in Section 2.2.1. The purification was performed as discussed in section 2.2.1 using Ni-affinity chromatography as a first purification step followed by gelfiltration. Both proteins were >95 % pure as judged from SDS-PAGE. For complex formation the individual proteins were mixed in a molar ratio of $CheY_6$: $CheA_3P1$ equals 2 : 1 and then incubated for 30min at 4°C. The protein complex was then further purified by gelfiltration. Complex formation was judged by SDS-PAGE and the corresponding fractions were pooled and concentrated for crystallization experiments.

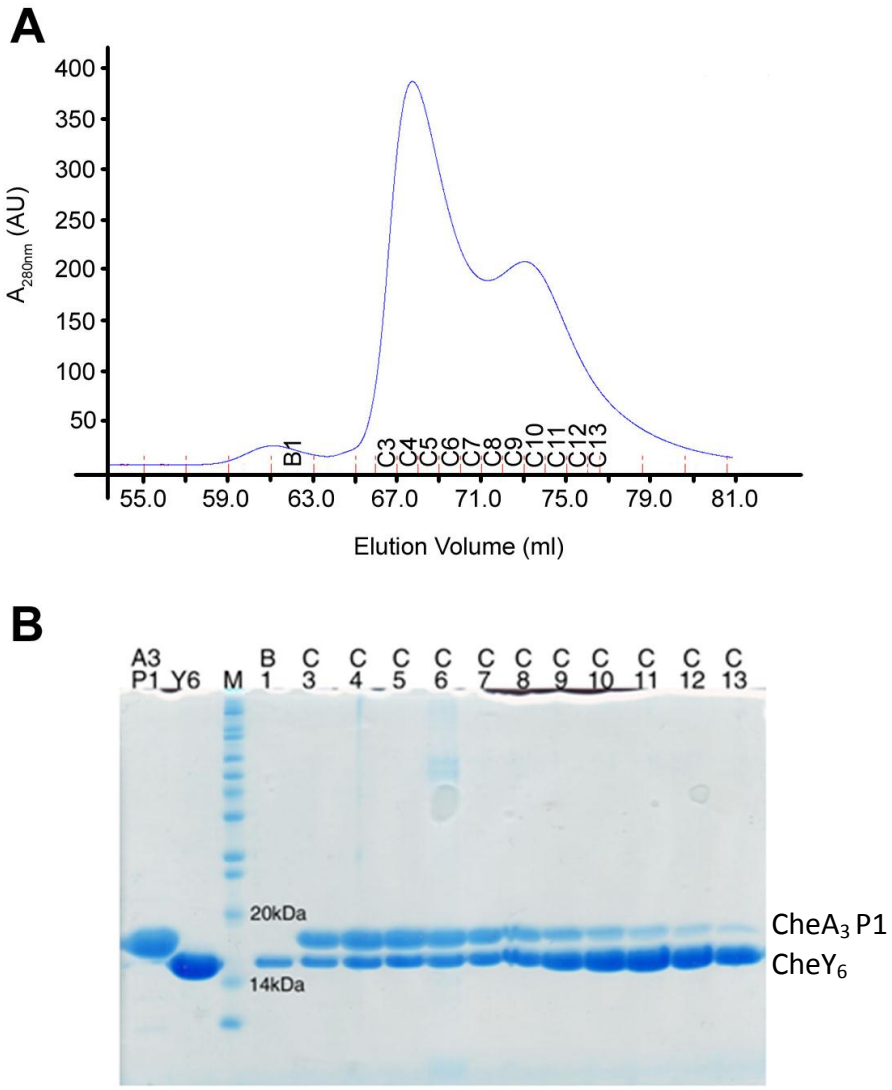


Figure 3.1: Purification of the $CheY_6$-$CheA_3P1$ complex. A) Gelfiltration profile from an S75 16/60 column for the protein complex after mixing and incubation on ice. B) SDS-PAGE gel of the fractions indicated in A). The individual proteins are shown as a control.

3.4 Crystallization and structure determination of the $CheY_6$-$CheA_3P1$ complex

Crystals of the $CheY_6$-$CheA_3P1$ complex could be obtained in 100 mM Bicine, pH 9, 1 M LiCl, 20 % PEG 6000. Crystals were optimized using a dilution series of the initial condition. Crystals reached their final size of 400x100x100 μm^3 after 12hrs. Diffraction data were

collected at 100 K with the crystals being flash-cooled in a cryo N_2 gas stream. Before flash-freezing, crystals were cryo-protected with perfluoropolyether oil PFOX125/03 (Lancaster Synthesis). The crystals diffracted to a resolution of 1.4 Å (Fig. 3.2). Due to the high resolution, several datasets with different high-resolution diffraction limits and exposure times were collected. This ensured a high level of completeness in the lowest as well as the highest resolution shells. The crystals belong to space group $P2_1$ with a solvent content of 43 %.

Attempts to solve the structure by molecular replacement using previously determined X-ray structures of Hpts or RRs did not result in reliable phases for the data. Therefore, $CheY_6$ was labelled with Seleno-Methionine (Se-Met), expressed in *E. coli* (see Section 2.2.1) and purified similar to the native protein. Crystals of the Se-Met labelled complex grew in the same condition and diffracted to a resolution of 2.3 Å.

The $CheA_3P1$-$CheY_6$ complex structure was determined by single-wavelength anomalous dispersion analysis. The positions of 8 selenium atoms were determined using SHELXD [145]. This solution was put into AUTOSHARP [121] for phase calculation, improvement and phase extension using the high resolution native data to 1.4 Å resolution. The resulting map was of high quality and allowed tracing of the whole polypeptide chain (Fig. 3.2). An initial model was built automatically using Arp/wARP [146] and manually adjusted using COOT[124]. The structure was refined using autoBUSTER[125], REFMAC[147] and Phenix[126].

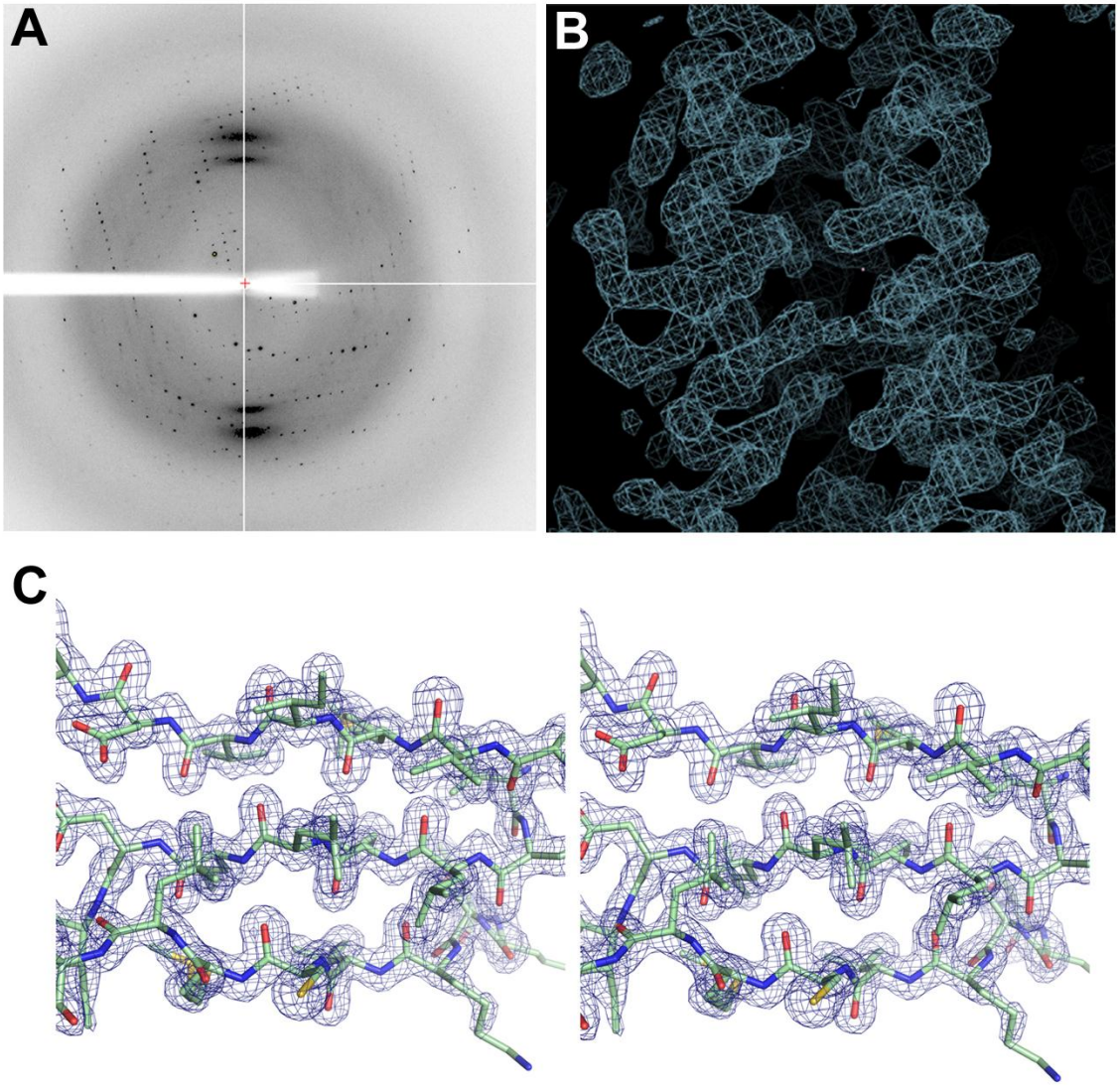


Figure 3.2: Data collection and structure solution of the $CheA_3P1$-$CheY_6$ complex. A) Representative diffraction image for the native, high-resolution dataset. B) Initial electron density map after phasing with AUTOSHARP [121] and solvent-flattening with Solomon as implemented in AUTOSHARP [121]. C) Representative $2F_{obs}$-F_{calc} map contoured at 1.5σ of the final, refined model.

In order to investigate the physiologically relevant, phosphorylated complex the structure of phosphorylated $CheA_3P1$ in complex with $CheY_6$ was solved as well. Since the rapid rate of phosphotransfer between $CheA_3P1$-P and $CheY_6$ does not allow crystallization of the wild-type, phosphorylated complex, a stable complex was formed by introducing two mutations,

D56N and S83A, in the active site of $CheY_6$. These mutations have previously been shown to abolish phosphotransfer from $CheA_3P1$-P [25]. The phosphorylated complex was still labile to hydrolysis therefore it was important to prepare the protein fresh before crystallization and to collect a dataset as soon as the crystals appeared. Despite best efforts the only dataset for this active, phosphorylated complex could be collected in-house on a Mar generator. Data was collected to 2.8 Å and the structure was solved by molecular replacement using Phaser [122] with the unphosphorylated structure as model.

The native dataset was collected at beamline I03 (lambda = 0.9757 Å) and the SeMet dataset at beamline I04 (lambda = 0.979 Å) at the Diamond Light Source, UK. Data for the phosphorylated complex were collected on a MAR345 detector(Marresearch) mounted on a Miromax007 generator (Rigaku), equipped with Varimax HR Osmic mirrors (Rigaku). X-ray data were processed and scaled with the HKL suite[119]. Data collection and refinement statistics can be found in Table 1.

Table 1. Data collection and refinement statistics

	SeMet	Unphosphorylated	Phosphorylated
Data collection			
Resolution (Å)	2.3 (2.38-2.30)[a]	1.40 (1.45-1.40)	2.80 (2.90-2.80)
Space group	$P2_1$	$P2_1$	P1
Cell dimension			
a, b, c (Å)	43.8, 62.7, 49.1	43.7, 62.0, 48.9	33.2, 43.4, 48.7
α, β, γ (°)	90, 101.1, 90	90, 101.3, 90	78.8, 86.6, 80.9
Redundancy	9.5 (10.1)	6.3 (3.7)	1.8 (1.8)
Completeness (%)	99.2 (99.0)	99.7 (98.9)	96.1 (94.5)
R_{merge} (%)	8.9 (19.7)	5.3 (14.8)	12.0 (61.6)
Avg I/σ	18.4 (12.9)	31.6 (9.8)	6.3 (1.2)
Refinement			
Resolution		38 - 1.40	42 - 2.80
No. Reflections		50287	6232
R_{work}/R_{free}		17.0/20.7	24.1[b]/28.4[b]
No. Atoms			
Protein		2040	1743
Water		272	-
B-factors (Å^2)			
Protein		28.5	86.5[c]
Water		38.9	-
r.m.s. Deviations			
Bond lengths (Å)		0.010	0.008
Bond angles (°)		1.220	0.950

[a] Numbers in parenthesis are for the highest resolution shell
[b] Rxpct, as described elsewhere [125]
[c] Standard deviation of ± 18 Å^2

3.5 Structure of the $CheA_3P1$-$CheY_6$ complex

The response regulator $CheY_6$ has the typical $(\alpha/\beta)_5$ topology seen for *E. coli* CheY [148] and other structurally characterized response regulators (Figure 3.3). There is a high degree of structural conservation when $CheY_6$ is compared to *E. coli* CheY with an rmsd of 1.4 Å over

115 Cα atoms (26 % sequence identity). Although the crystallization condition as well as all the buffers used for purification included Mn^{2+}, no additional density in the conserved binding site for divalent cations can be observed. However, the conformation of the metal coordinating residues Asp56, Asp9 and Asp10 is most similar to that found in the structure of Mg^{2+} bound CheY [149]. The only difference is the backbone carbonyl of Glu58 which is facing away rather than pointing towards the potential metal binding site (Figure 3.4). Remarkably, $CheY_6$ has an elongated loop region connecting β5 and α5 (Figure 3.3). This loop comprises 13 residues (res. 107-119) in $CheY_6$ compared to only three residues (109-111) in *E. coli* CheY. In the crystal, this loop is only partially ordered suggesting that it is highly flexible. Residues 113-118 could not be traced and were omitted from the final model. The N-terminal region of this β5 -α5 loop in combination with α1 of $CheY_6$ form the vast majority of contacts to $CheA_3P1$ (Figure 1).

$CheA_3P1$ forms a four-helix bundle (αA-αD) with an additional C-terminal helix (αE) (Figure 1). Despite low sequence identity it is structurally very similar to previously determined CheA P1 structures [150-152]. Comparison with CheA P1 from *Salmonella enterica* serovar Typhimurium [151] and *Thermotoga maritima* [152] gives an rmsd of 1.2Å over 116 Cα atoms (21% sequence identity) and 1.5Å over 101 Cα (18% sequence identity), respectively. The site of phosphorylation, His51 in $CheA_3P1$, is located on αB, in close proximity to the active site of the response regulator (Figure 3.3) with the phosphoacceptor Asp56 on $CheY_6$ being 7.5Å from His51 on $CheA_3P1$. Helices αA and αB face the response regulator and together with the αB-αC loop form the interface with $CheY_6$ (Figure 3.3).

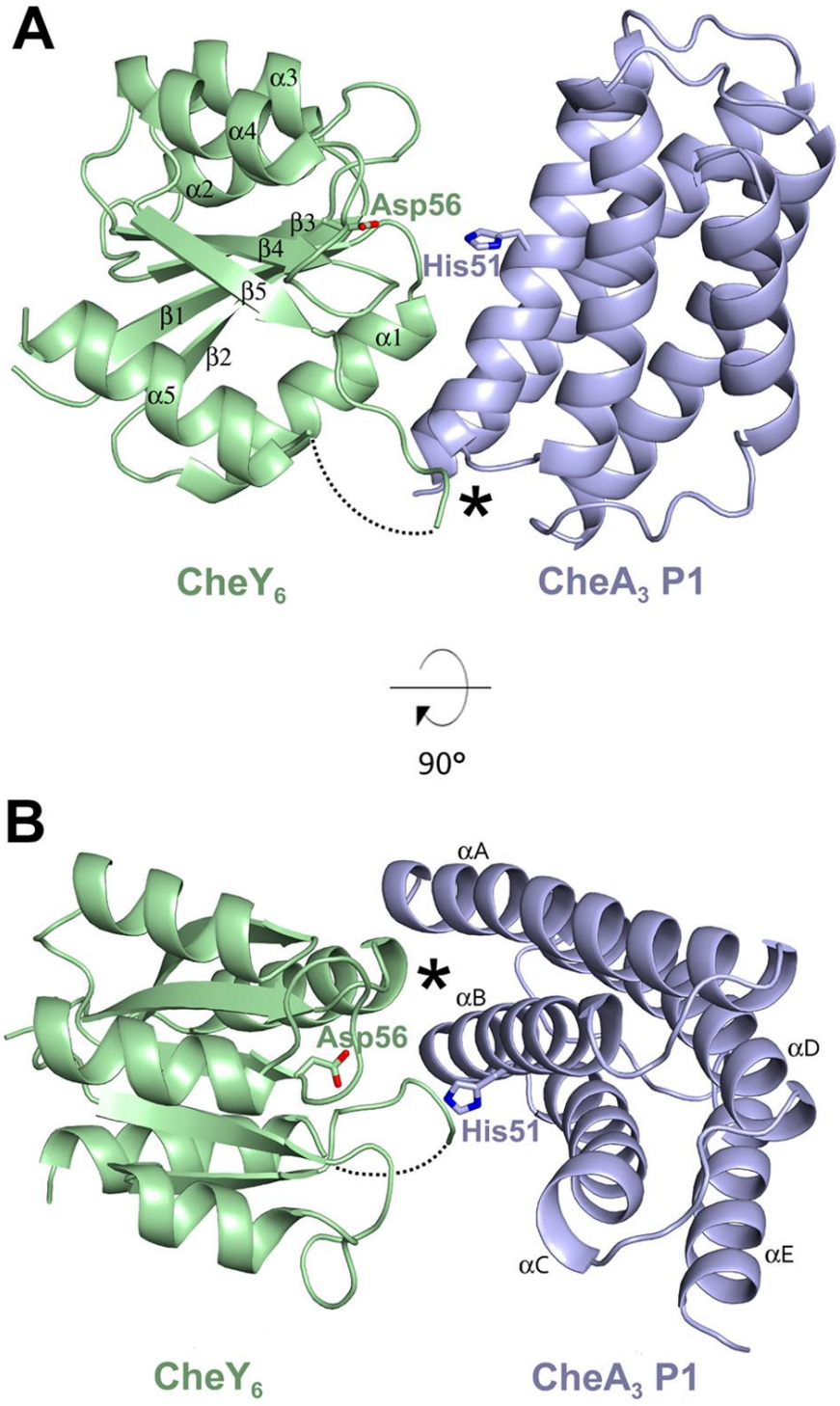


Figure 3.3: Structure of the $CheA_3P1$-$CheY_6$ complex. (A-B) $CheA_3P1$ is coloured in lightblue whereas $CheY_6$ is coloured in palegreen. The phosphorylatable residue of $CheA_3P1$, His51, and the phosphoacceptor on $CheY_6$, Asp56, are shown in stick representation. Secondary structure elements are labelled in black. Residues 113-118 of $CheY_6$ could not be traced and are depicted as dotted line. (A) Sideview onto the $CheA_3P1$-$CheY_6$ complex with the interaction between β5 and α5 in $CheY_6$ and αB and the following loop connecting αB and αC on $CheA_3P1$ marked with an asterisk. (B) Topview showing the major site of interaction between the N-terminal end of α1 on $CheY_6$ and αA/αB on $CheA_3P1$ marked by an asterisk.

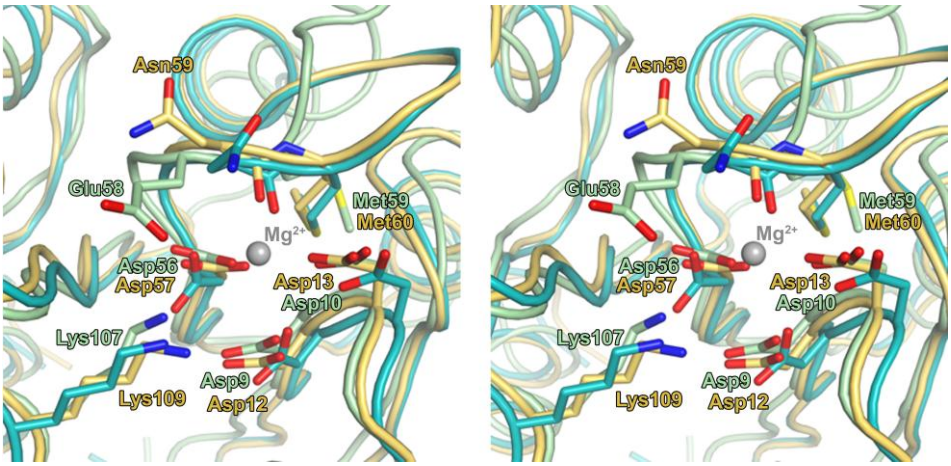


Figure 3.4: Stereoview of a superposition of the metal binding site of $CheY_6$ and *E. coli* CheY. $CheY_6$ is coloured in palegreen, *E. coli* CheY with Mg^{2+} bound (PDB-code: 1CHN) in yellow and *E. coli* CheY without Mg^{2+} (3CHY) in teal. Structures were aligned on their secondary structure elements using SSM implemented in COOT [124]. Residues involved in the coordination of Mg^{2+} in *E. coli* CheY are shown in stick representation for all three structures. Residues in *E. coli* structures are only labelled for the Mg^{2+} bound form. Mg^{2+} from the *E. coli* structure is shown as grey sphere. $CheY_6$ resembles the Mg^{2+} bound form of *E. coli* CheY despite not having a divalent cation bound to its metal binding site.

3.6 Structure of phosphorylated $CheA_3P1$ in complex with $CheY_6$ Asp56Asn,Ser83Ala

The phosphorylated structure shows clear additional density adjacent to Nε2 of His51 of $CheA_3P1$ in agreement with phosphorylation of this residue (Figure 3.5). Due to the moderate resolution of the structure, residues 60-65, 85-97 and 111-121 on $CheY_6$ could not be traced and were not included in the final model. Compared to the unphosphorylated complex, $CheA_3P1$-P undergoes a rigid body translation of 2.1 Å relative to the response regulator (Figure 3.6A). This realignment of $CheA_3P1$-P positions the phosphorylated His51 closer to the phosphoacceptor on $CheY_6$, Asp56 (Figure 3.6B).

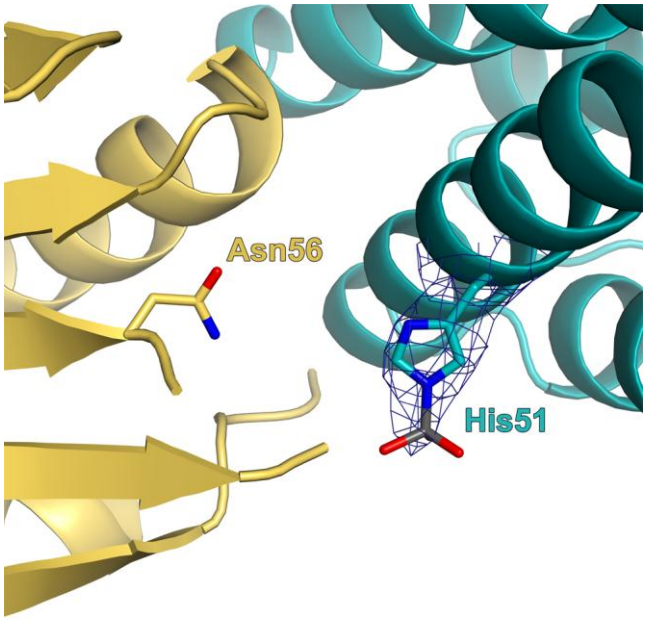


Figure 3.5: Electron density of phosphorylated His51 in $CheA_3P1$ in the active complex structure. The orientation is similar to Fig. 2b. The density represents a $2F_{obs}$-F_{calc} map contoured at 1.5 σ and calculated after initial rigid body refinement in Phaser [122] and one round of positional refinement in autoBUSTER [125], both omitting the phosphate group from the model.

The phosphorylated His51 is facing away from the active site of $CheY_6$, however, a 180° flip of this sidechain would put it in near linear geometry and place the phosphoryl group within 4.5 Å of Asp56 of $CheY_6$ (Figure 3.6B). The orientation of the phosphorylated His51 seen in the crystal structure is likely caused by the lack of a divalent cation in the metal binding site of the response regulator leading to electrostatic repulsion from the $CheY_6$ active site.

The interface seen in the phosphorylated complex (total buried surface area of 605 $Å^2$) is slightly more extensive than in the unphosphorylated complex (530 $Å^2$), however, the contributions of specific residues involved in the interaction are almost equal. The only difference is a hydrogen bond formed between Glu58 of CheY6 and His51 of CheA3P1 in the unphosphorylated complex which is released in the phosphorylated complex. As this is not contributing to the interface of the physiologically important complex it is therefore not discussed further. Because the interfaces in the phosphorylated and unphosphorylated

complex are highly similar, the following discussion of the interactions between $CheA_3P1$ and $CheY_6$ are based on the high-resolution, unphosphorylated structure.

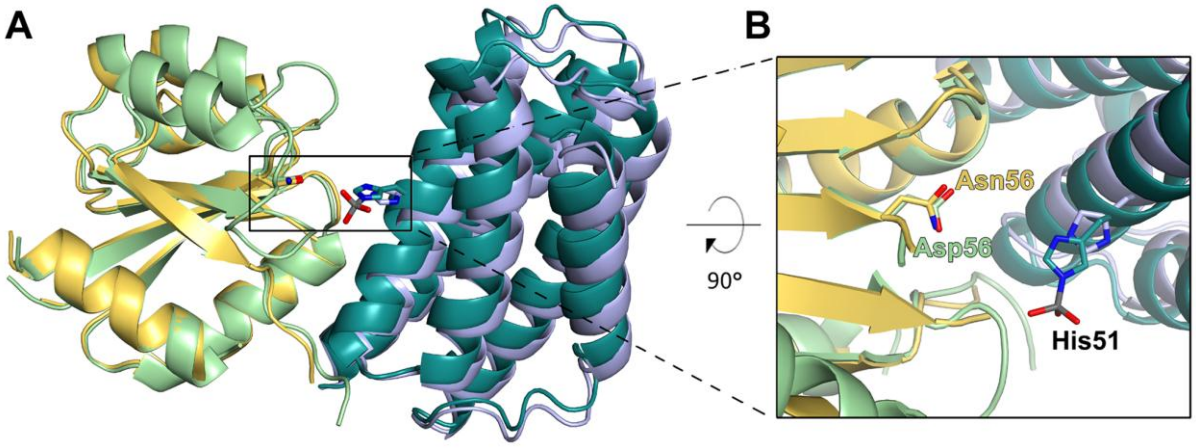


Figure 3.6: Superposition of the $CheA_3P1$-$CheY_6$ complex in the phosphorylated and unphosphorylated conformations. (A-B) Structures were aligned onto the response regulator. Colour coding for the unphosphorylated conformation as in Figure 3.3, for the phosphorylated conformation $CheY_6$ is coloured in yellow and $CheA_3P1$ in teal. The active site residues His51 ($CheA_3P1$) and Asp56/Asn56 ($CheY_6$ unphosphorylated/phosphorylated complex conformation) are shown in stick representation. (A) Overview highlighting the 2.1 Å rigid body translation of $CheA_3P1$ towards $CheY_6$. (B) Close-up view onto the active site showing the movement of His51 towards Asp56/Asn56.

3.7 The interface between $CheA_3P1$ and $CheY_6$

The interface between $CheA_3P1$ and $CheY_6$ is dominated by hydrophobic interactions with only one hydrogen bond and no salt bridges formed between the two proteins. The buried surface area of the interface is small (530 $Å^2$) indicative of a weak interaction, consistent with the transient nature of the complexes [153]. A continuous interface is formed by two sites of interaction complemented by a hydrogen bond formed between Ser83 on $CheY_6$ and Arg58 on $CheA_3P1$. One of these sites lies between the elongated loop region between β5

and α5 in $CheY_6$ and αB and the following loop connecting αB and αC on $CheA_3P1$ (marked with an asterisk in Figure 3.3A). The other site lies between the N-terminal end of α1 on $CheY_6$ and αA/αB on $CheA_3P1$ (marked with an asterisk in Figure 3.3B). The latter is the major binding site including ~70 % of the total buried surface area (358 Å2).

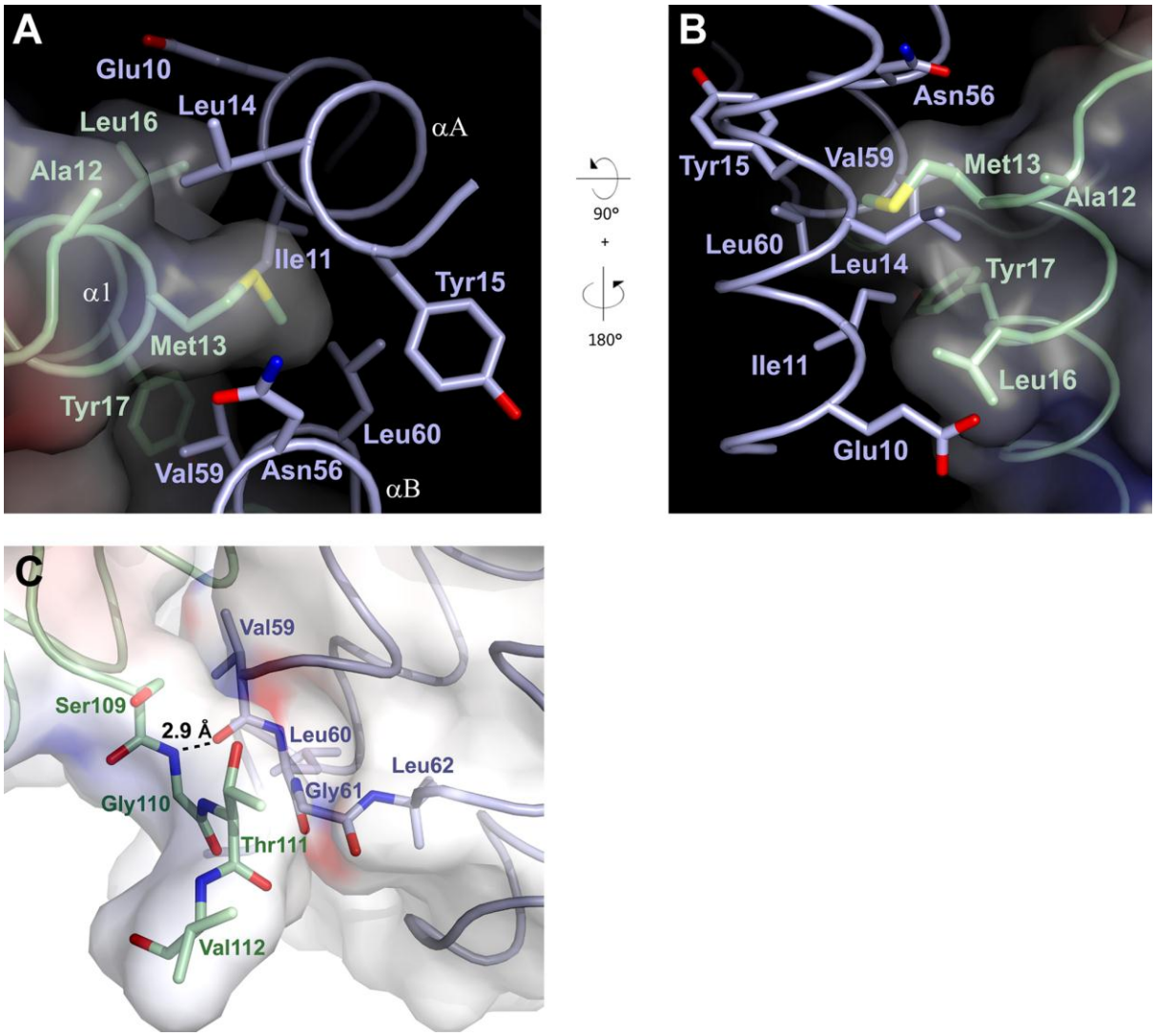


Figure 3.7: Close-up view of the two binding sites between $CheA_3P1$ and $CheY_6$. (A-C) Colour coding as in Figure 3.3. Residues that are involved in the interaction are shown in stick representation. (A) Detailed view onto the major binding site between $CheY_6$ and $CheA_3P1$. Orientation as in Figure 3.3B, residues 19-53 of CheA3 P1 are omitted for clarity. (B) Solvent accessible surface of $CheY_6$ colored by electrostatic potential contoured at ± 10kT. Met13 fits snugly into the hydrophobic pocket on $CheA_3P1$. (C) Close-up view onto the second binding site. Orientation as in Figure 3.3A. The solvent-accessible surface area is coloured by electrostatic potential contoured at ± 10kT. The hydrogen bond between Gly110 of $CheY_6$ and Val59 of $CheA_3P1$ is marked with a dotted line.

In this region helices αA and αB of $CheA_3P1$ form a hydrophobic pocket constituted of Ile11, Leu14 and Tyr15 on αA and Asn56, Val59 and Leu60 on αB (Figures 3.7A and 3.7B). This pocket is situated adjacent to α1 on $CheY_6$ and in ideal position to accommodate Met13 on the N-terminal region of this helix. This Met-finger is protruding from α1 and its side chain fits snugly into the hydrophobic pocket, with 99.7 % of its accessible surface area being buried (Figure 3.7B). The hydrophobic interaction is complemented by Ala12, Leu16 and Tyr17 of α1 on $CheY_6$, all facing the same side on $CheA_3P1$ as Met13 (Figure 3.7A). Ala12 of $CheY_6$ extends the hydrophobic interface by interaction with Leu14 of $CheA_3P1$ whereas Leu16 of $CheY_6$ interacts with Ile11 and the aliphatic part of Glu10 on αA of $CheA_3P1$. Tyr17 extends the binding surface towards the β5-α5 loop by interacting with Val59 at the C-terminus of αB. The elongated loop between β5 and α5 of $CheY_6$ forms van der Waals interactions with Ser109, Gly110 and Thr111 of $CheY_6$ contacting Ile59, Gly61 and Ser63 on αB and the following loop region on $CheA_3P1$ (Figure 3.7C). A main chain hydrogen bond is formed between Gly110 and Val59. Of these interactions, the Met13 finger has the largest contribution towards the binding interface on $CheY_6$, accounting for almost a third (156 $Å^2$) of the total buried surface area. Together with Ala12, Leu16 and Tyr17 it is accounting for over 60 % (321 $Å^2$) of the interface. On $CheA_3P1$, the residues involved in the interaction with $CheY_6$ are slightly less clustered and spatially further apart. Yet the three residues with the biggest contribution towards the total buried surface area Val59 (103 $Å^2$, 20% of total buried surface area), Leu14 (64 $Å^2$, 12%) and Ile11 (48 $Å^2$, 9%) together account for almost half of the overall binding surface. This analysis shows that only a small number of residues are necessary to form the majority of interactions within the complex. On $CheY_6$ these are

all located on α1. Within this helix, Met13 has the most crucial role in the hydrophobic interaction. A sequence alignment of all identified response regulators in *R. sphaeroides* shows that only $CheY_6$ and $CheB_2$ have a Met residue at position 13 whereas all others either have a Ser, Thr or Ala at this position (Figure 3.8). As $CheA_3$-P serves as a phosphodonor for both $CheY_6$ and $CheB_2$ this suggests that Met13 has an important role in determining the specificity for binding to CheA3P1.

β1
α1
β2
α2
β3
α3
β4
α4
β5
α5
CheY6
CheB1
CheB2
ECCheB
CheY1
CheY2
CheY3
CheY4
CheY5
ECCheY

Figure 3.8: Structure guided sequence alignment of the chemotaxis response regulators from E. coli and R. sphaeroides. Alignment is based on the structures of E. coli CheY (PDB-code: 3CHY), E. coli CheB (1A2O), R. sphaeroides CheY3 (see Chapter 4) and R. sphaeroides CheY6. Secondary structure is shown for CheY6. Residues involved in binding of CheY6 to CheA3P1 are marked with a star (contributes > 30% to total buried surface area), square (>15%) or circle (>5%).

3.8 Biophysical and functional characterization of the interface

The structural characterization lead to the conclusion that residues 11-17 of $CheY_6$ collectively account for ~67 % of its buried surface area in the $CheA_3P1$-$CheY_6$ complex, with Met13, Leu16 and Tyr17 each contributing ~30 %, ~16 % and ~8 %, respectively. To confirm that this surface of $CheY_6$ is involved in the interactions with $CheA_3P1$-P that lead to phosphotransfer in solution, the residues at this surface were mutated to mimic those found in the non-cognate response regulators, $CheY_3$ and $CheY_4$. Met13 was changed to Ser, and Tyr17 was changed to Met as found in both $CheY_3$ and $CheY_4$. $CheY_3$ and $CheY_4$ both have Cys at position 16, however it has previously been shown that $CheY_3$ Cys16Ser and $CheY_4$ Cys16Ser behave indistinguishably from their native proteins in phosphotransfer assays (Steven L. Porter, personal communication), and therefore to avoid any potential problems from disulphide bond formation during the purification Leu16 was changed to Ser rather than Cys. Two mutant proteins were expressed and purified, $CheY_6$ Met13Ser and $CheY_6$ Met13Ser, Leu16Ser, Tyr17Met.

Surface plasmon resonance (SPR) binding assays showed that wild-type $CheY_6$ binds to $CheA_3P1$ with an affinity of 218 µM (Figure 3.9). This weak interaction is in agreement with the transient nature of the complex and the small buried surface area between the two

proteins as observed in the crystal structure. Both the single mutant, CheY6 Met13Ser, and the triple mutant CheY6 Met13Ser, Leu16Ser, Tyr17Met showed a remarkable decrease in affinity with a K_d of more than 1 mM (Figure 3.9 and Figure 3.10).

A

Interaction	K_d [μM]
$CheA_3$ P1 - $CheY_6$ wt	218 ± 8
$CheA_3$ P1 - $CheY_6$(M13S)	> 1000
$CheA_3$ P1 - $CheY_6$(M13S,L16S,Y17M)	> 1000
$CheA_3$ P1 - $CheY_4$ wt	ND
$CheA_3$ P1 - $CheY_4$(P12A,S13M)	678 ± 34
$CheA_3$ P1 - $CheY_4$(P12A,S13M,C16L,M17Y)	365 ± 35

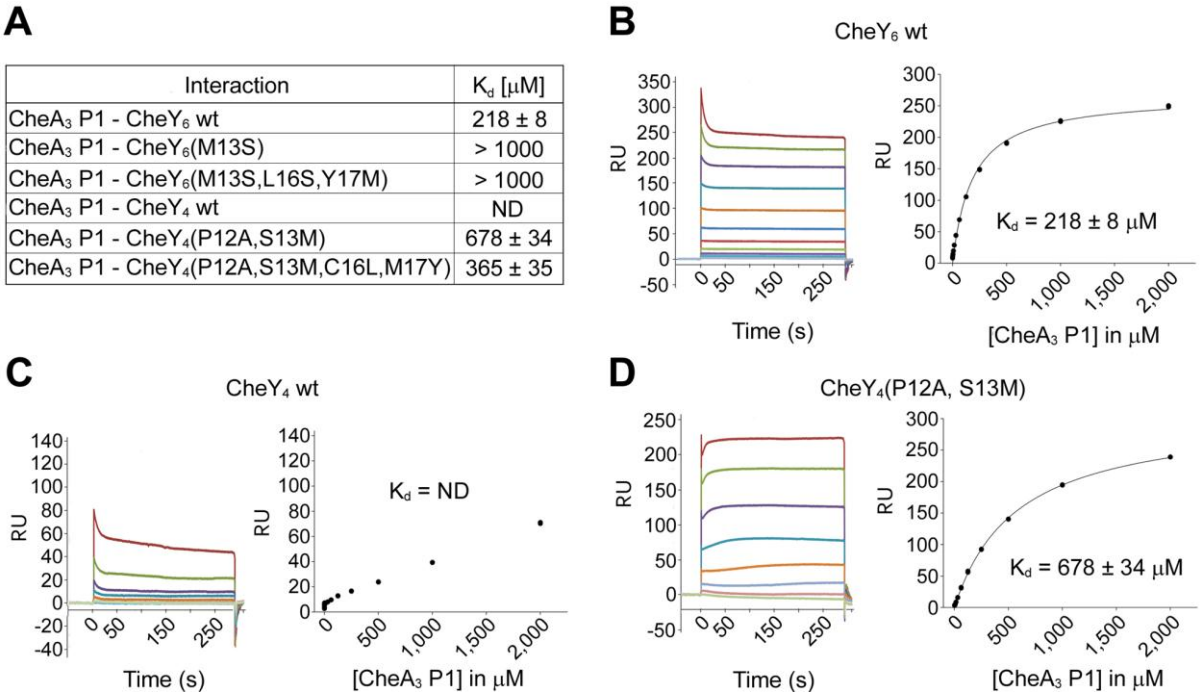


Figure 3.9: Binding of $CheA_3P1$ to the response regulators. (A) Table of binding constants (K_d) measured by SPR between $CheA_3P1$ and $CheY_4$, $CheY_6$ and their mutant versions. Data are expressed as mean ± s.e.m. ND, not determinable. (B-D) Binding of $CheA_3P1$ to $CheY_6$ wt, $CheY_4$ wt and $CheY_4$ Pro12Ala, Ser13Met). Left, representative sets of experimental sensorgrams from typical equilibrium-based binding experiments, with reference subtraction. Different concentrations of $CheA_3P1$ were injected over surfaces coupled with the respective response regulator. For all injections, the experimental traces reached equilibrium and returned to baseline after the injection. Right, plot of the equilibrium binding response (response units (RU)) against $CheA_3P1$ concentration ranging from 120 nM to 2 mM. Within one experiment each concentration was measured twice. All experiments were performed in duplicate. Best-fit binding curves corresponding with a 1:1 binding model are shown as lines.

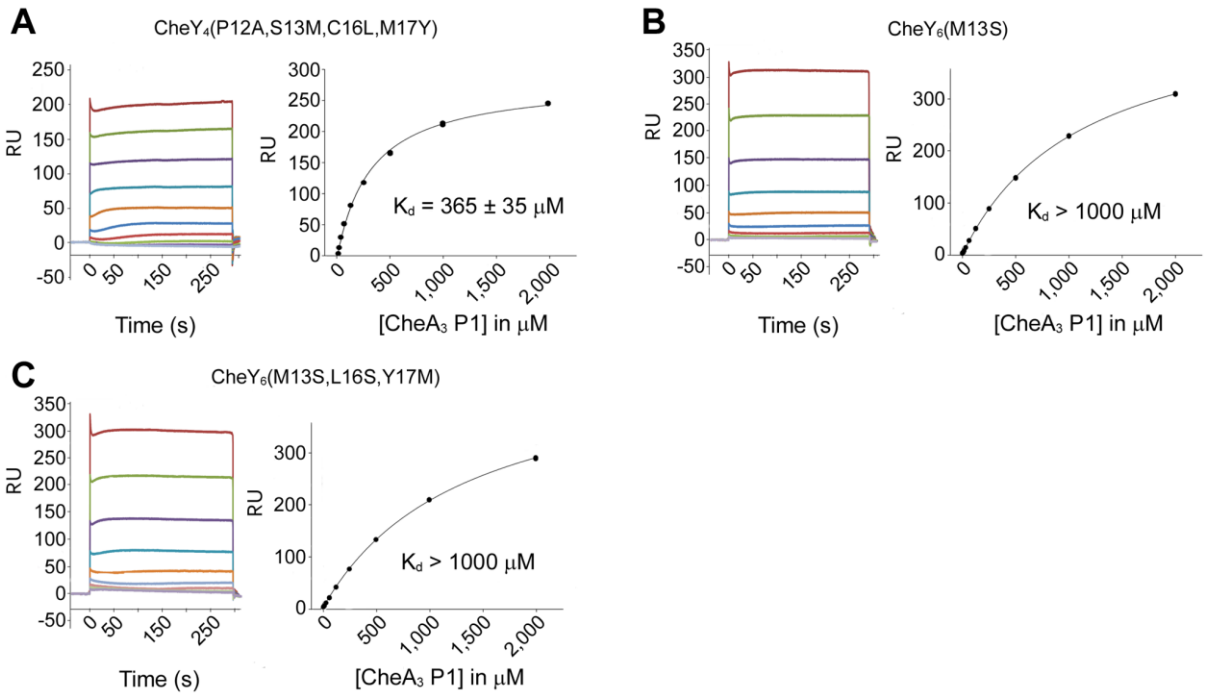


Figure 3.10: Binding of CheA$_3$P1 to mutants of CheY$_4$ and CheY$_6$. (A-C) Binding of CheA$_3$P1 to CheY$_4$ Pro12Ala, Ser13Met, Cys16Leu, Met17Tyr, CheY$_6$ Met13Ser and CheY$_6$ Met13Ser,Leu16Ser,Tyr17Met. Left, representative sets of experimental sensorgrams from typical equilibrium-based binding experiments, with reference subtraction. Different concentrations of CheA3P1 were injected over surfaces coupled with the respective response regulator. For all injections, the experimental traces reached equilibrium and returned to baseline after the injection. Right, plot of the equilibrium binding response (response units (RU)) against CheA$_3$P1 concentration ranging from 120 nM to 2 mM. Within one experiment each concentration was measured twice. All experiments were performed in duplicate. Best-fit binding curves corresponding with a 1:1 binding model are shown as lines. Experiments with CheY$_6$ Met13Ser and CheY$_6$ Met13Ser,Leu16Ser,Tyr17Met did not reach saturation due to very low affinity. Thus the K_d value is estimated.

In addition to the SPR binding studies the interface was also characterized in a functional context using a radioactive phosphotransfer assay. The assay itself was carried out by Steven L. Porter using constructs and proteins that I had prepared.

Consistent with the SPR binding assays, the rate of phosphotransfer from CheA$_3$P1-P to

each of the two mutant $CheY_6$ proteins as measured by ^{32}P phosphotransfer from $CheA_3$-P was much slower than to wild-type $CheY_6$ (Figure 3.11A-C), with phosphotransfer to the triple mutant, $CheY_6$(M13S,L16S,Y17M) being slowest. These results underline the conclusions drawn from the structural characterization of the complex and stress the importance of the Met finger at position 13 for binding of $CheY_6$ to $CheA_3P1$. The hydrophobic interaction mediated by Leu16 and Tyr17 also adds significantly to recognition of $CheY_6$ by $CheA_3P1$.

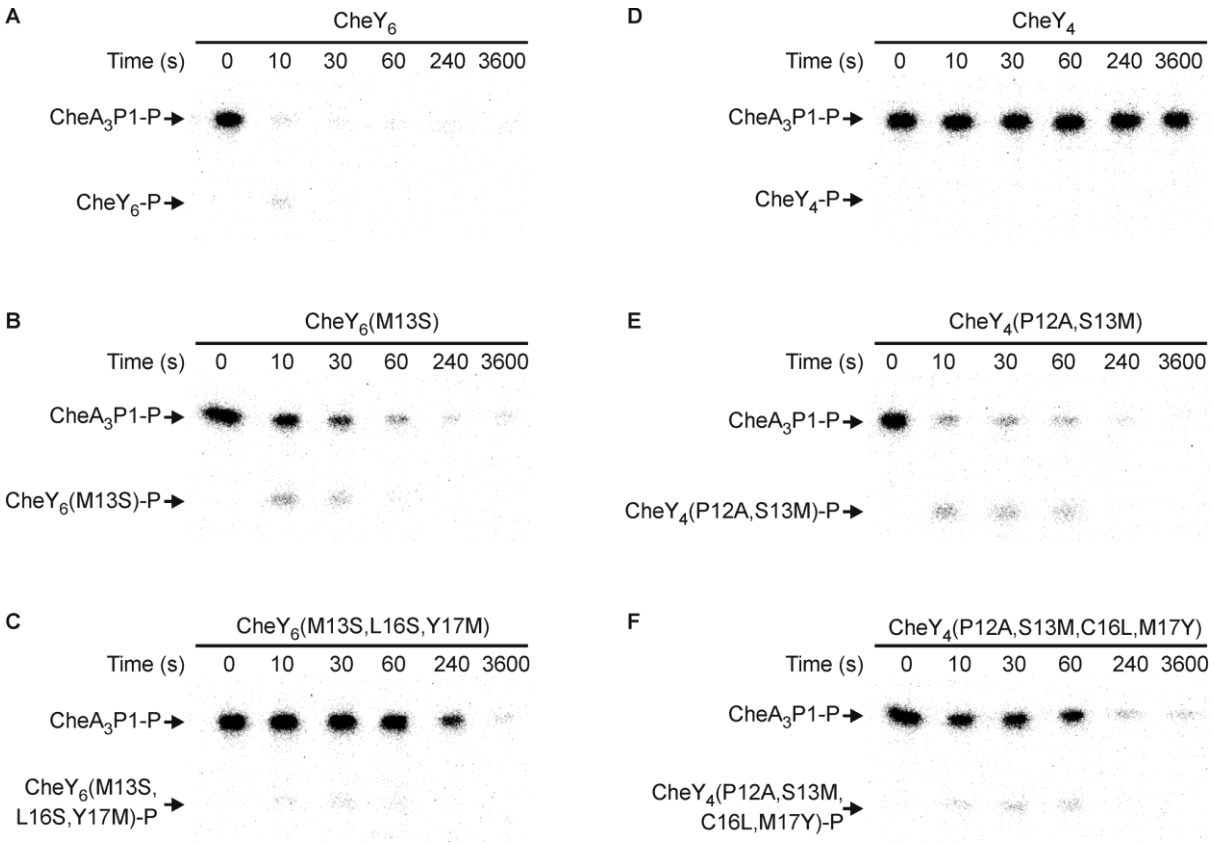


Figure 3.11: Rewiring the phosphotransfer reactions. Phosphorimages of SDS-PAGE gels measuring phosphotransfer from $CheA_3P1$-P to (A) $CheY_6$, (B) $CheY_6$ Met13Ser, (C) $CheY_6$ Met13Ser,Leu16Ser,Tyr17Met, (D) $CheY_4$, (E) $CheY_4$ Pro12Ala,Ser13Met, (F) $CheY_4$ Pro12Ala,Ser13Met,Cys16Leu,Met17Tyr). 5 μM of CheY was added to 1 μM of $CheA_3P1$-^{32}P. 10 μl reaction samples were taken at the time points indicated and quenched in 20 μl of 1.5 X SDS/EDTA loading dye. The quenched samples were analyzed by SDS-PAGE and detected by phosphorimaging. The assay was carried out by Steven L. Porter.

3.9 Rewiring chemotaxis

After successfully identifying the determinants for specificity in the $CheA_3P1$-$CheY_6$ complex, this information was used to design novel, synthetic two-component pairs with non-cognate response regulators and $CheA_3P1$ as the kinase.

The first non-cognate response regulator used for this was $CheY_4$. The sequence of $CheY_4$ is 36 % identical to $CheY_6$, however, unlike $CheY_6$, $CheY_4$ is not a cognate response regulator of $CheA_3P1$-P (Figure 3.11D). Consistent with this, SPR assays failed to detect a significant interaction between $CheA_3P1$ and $CheY_4$ (Figure 3.9C). In an attempt to rewire the phosphotransfer specificity of $CheY_4$ to allow phosphotransfer from $CheA_3P1$-P, we substituted the $CheY_4$ residues corresponding to the positions shown to interact in the $CheA_3P1$-$CheY_6$ structure, so that they matched $CheY_6$. Two mutant proteins were produced: $CheY_4$ Pro12Ala,Ser13Met and $CheY_4$ Pro12Ala,Ser13Met,Cys16Leu,Met17Tyr. Ala12 was included in both mutant proteins since $CheY_4$ has a Pro at position 12 that might influence the orientation of α1, and thus interfere with the proper positioning of Met13 for interaction with the hydrophobic pocket. SPR assays showed that both mutant proteins bind $CheA_3P1$ (Figure 3.9 and Figure 3.10) and in phosphotransfer assays, both mutant proteins were phosphorylated by $CheA_3P1$-P (Figure 3.11E-F). Phosphotransfer was fastest from $CheA_3P1$-P to $CheY_4$ Pro12Ala,Ser13Met, with most of the initial $CheA_3P1$-P dephosphorylated within 10 seconds (Figure 3.11E).

Interestingly, while the rate of phosphotransfer from $CheA_3P1$-P to $CheY_4$ Pro12Ala,Ser13Met was faster than to $CheY_4$ Pro12Ala,Ser13Met,Cys16Leu,Met17Tyr, the SPR results show that $CheA_3P1$ interacts slightly more strongly with

$CheY_4$ Pro12Ala,Ser13Met,Cys16Leu,Met17Tyr than with $CheY_4$ Pro12Ala,Ser13Met (Figure 3.9 and 3.10). This apparent discrepancy could be explained by the alignment of the phosphorylated histidine and the phosphorylatable aspartate in the $CheY_4$ Pro12Ala,Ser13Met,Cys16Leu,Met17Tyr-$CheA_3P1$-P complex which might be slightly less optimal for catalysis than in the $CheY_4$ Pro12Ala,Ser13Met-$CheA_3P1$-P complex. Nevertheless, both methods show that mutating $CheY_4$ so that it resembles $CheY_6$ at the contact sites with $CheA_3P1$, enhances both binding affinity and phosphotransfer rate. These results demonstrate that substitution of just two residues is sufficient to rewire the phosphotransfer specificity of $CheY_4$.

Having successfully reengineered the phosphotransfer specificity of $CheY_4$ by introducing Ala12 and Met13, the same approach was used to rewire the specificity of $CheY_1$, $CheY_3$, $CheY_5$ and *E. coli* CheY (Figure 3.12). These proteins share between 30 % and 33 % sequence identity with $CheY_6$. In all cases, the mutant proteins containing the alanine and methionine substitutions were phosphorylated more rapidly by $CheA_3P1$-P than their wild-type counterparts. Although $CheY_1$ and *E. coli* CheY both lack a methionine residue at the position corresponding to M13 of $CheY_6$ they are both phosphorylated by $CheA_3P1$-P (Figure 3.12A and 3.12C); however, phosphotransfer to their corresponding alanine and methionine substitution mutant proteins proceeded even faster, with almost complete dephosphorylation of $CheA_3P1$-P within 10 s (Figure 3.12B and 3.12D). Similar to $CheY_4$, wild-type $CheY_3$ and $CheY_5$ were not phosphorylated by $CheA_3P1$-P (Figure 3.12E and 3.12G) and the effect of the alanine and methionine substitutions was to allow rapid phosphotransfer from $CheA_3P1$-P (Figure 3.12F and 3.12H). These results demonstrate that

the response regulator residues at the positions equivalent to Ala12 and Met13 of $CheY_6$ play a major role in determining phosphotransfer specificity.

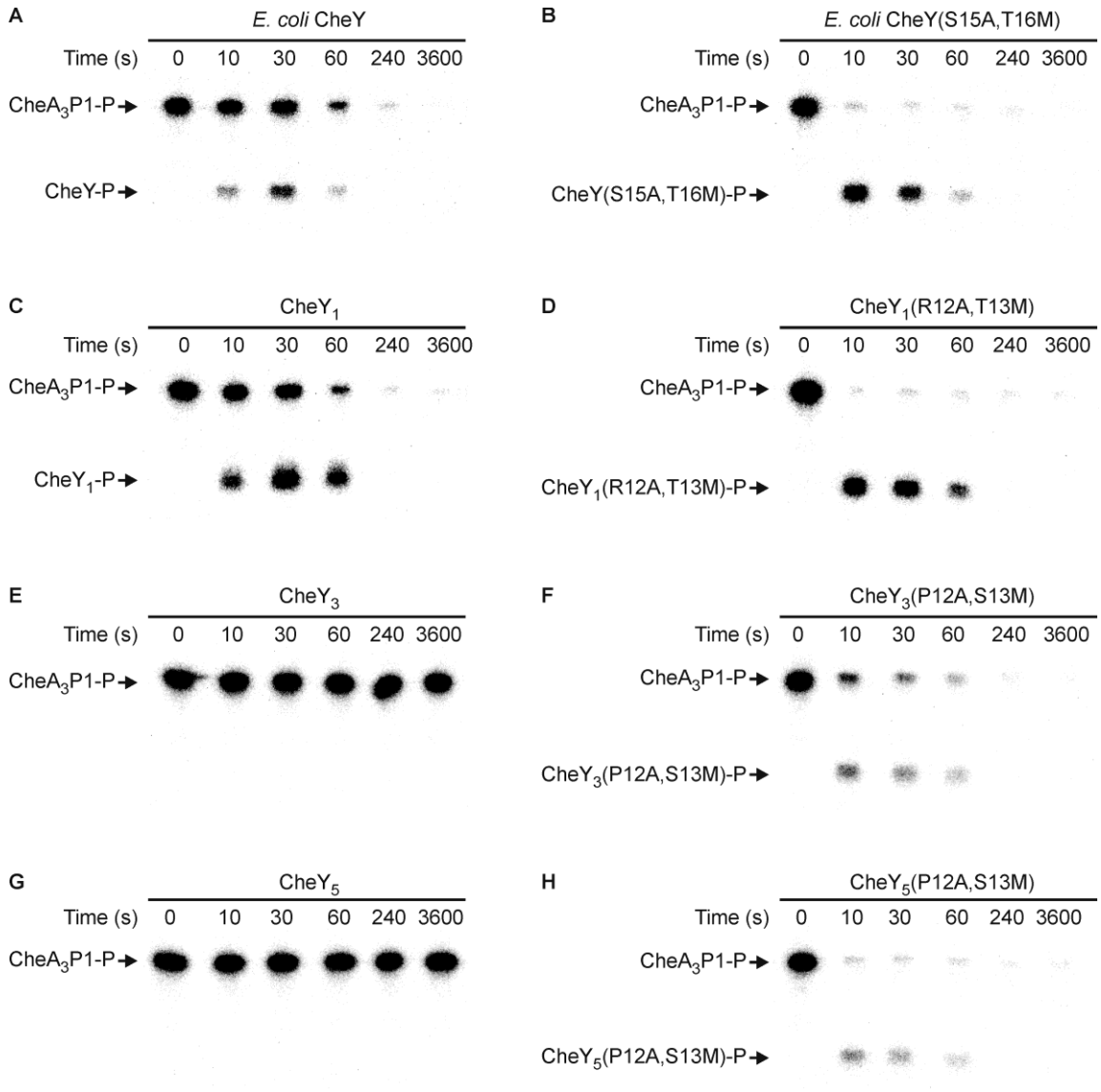


Figure 3.12: Rewiring phosphotransfer reactions involving other CheYs by introduction of Ala12 and Met13. Phosphorimages of SDS-PAGE gels measuring phosphotransfer from $CheA_3P1$-P to (A) *E. coli* CheY, (B) *E. coli* CheY Ser15Ala,Thr16Met, (C) $CheY_1$, (D) $CheY_1$ Arg12Ala,Thr13Met, (E) $CheY_3$, (F) $CheY_3$ Pro12Ala,Ser13Met, (G) $CheY_5$, (H) $CheY_5$ Pro12Ala,Ser13Met. 5 µM of CheY was added to 1 µM of $CheA_3P1$-^{32}P. 10 µl reaction samples were taken at the time points indicated and quenched in 20 µl of 1.5 X SDS/EDTA loading dye. The quenched samples were analyzed by SDS-PAGE and detected by phosphorimaging. The assay was carried out by Steven L. Porter.

3.10 Discussion

Two-component signaling systems depend on a high level of selectivity between histidine kinases and their cognate response regulators to prevent cross-talk between different pathways. Here, structural and functional data are presented that elucidate how this specificity is mediated on a molecular level. The key residues for molecular recognition were identified and allowed rewiring the pathway. Considering the high level of structural homology within the receiver and Hpt domain families, the presented structure together with complementary information, e.g. analysis of amino-acid covariation [154-156], is likely to facilitate the rational reengineering of CheA-like histidine kinase–response regulator pairs for use, for example, in synthetic signalling circuits.

The structure of $CheA_3P1$ in complex with $CheY_6$ is the first structure of a Hpt domain of a CheA protein in complex with a response regulator. To date there are only two other structures of a Hpt domain in complex with a response regulator, namely Spo0B/Spo0F [157,158] from *B. subtilis* and YPD1/SLN1 [159,160] from *S. cerevisiae*. While the overall structure of the response regulator is well conserved in all three complexes, there are remarkable differences in the Hpt domains. Spo0B exists as a dimer in which the two alpha helices of the helical hairpin domain of each protomer associate to form a four helix bundle. In addition, Spo0B has a C-terminal domain with an α/β fold, which is involved in binding of the response regulator. YPD1 is a monomeric Hpt protein consisting of a four helix bundle with an additional short helix at the N-terminus. This N-terminal helix is involved in response regulator binding and complements the interface formed by helices αB and αC. When the three complexes were superimposed with respect to the receiver domains, $CheA_3P1$ aligned

well with YPD1 but only poorly with Spo0B (Figure 3.13). However, the distance between the phosphoacceptor on $CheY_6$, Asp56, and the phosphodonor on $CheA_3P1$, His51, is 7.5 Å in the unphosphorylated structure and thus considerably larger than in the YPD1/SLN1 complex (4.5 Å), possibly suggesting differences in the mechanism of phosphoryl group transfer.

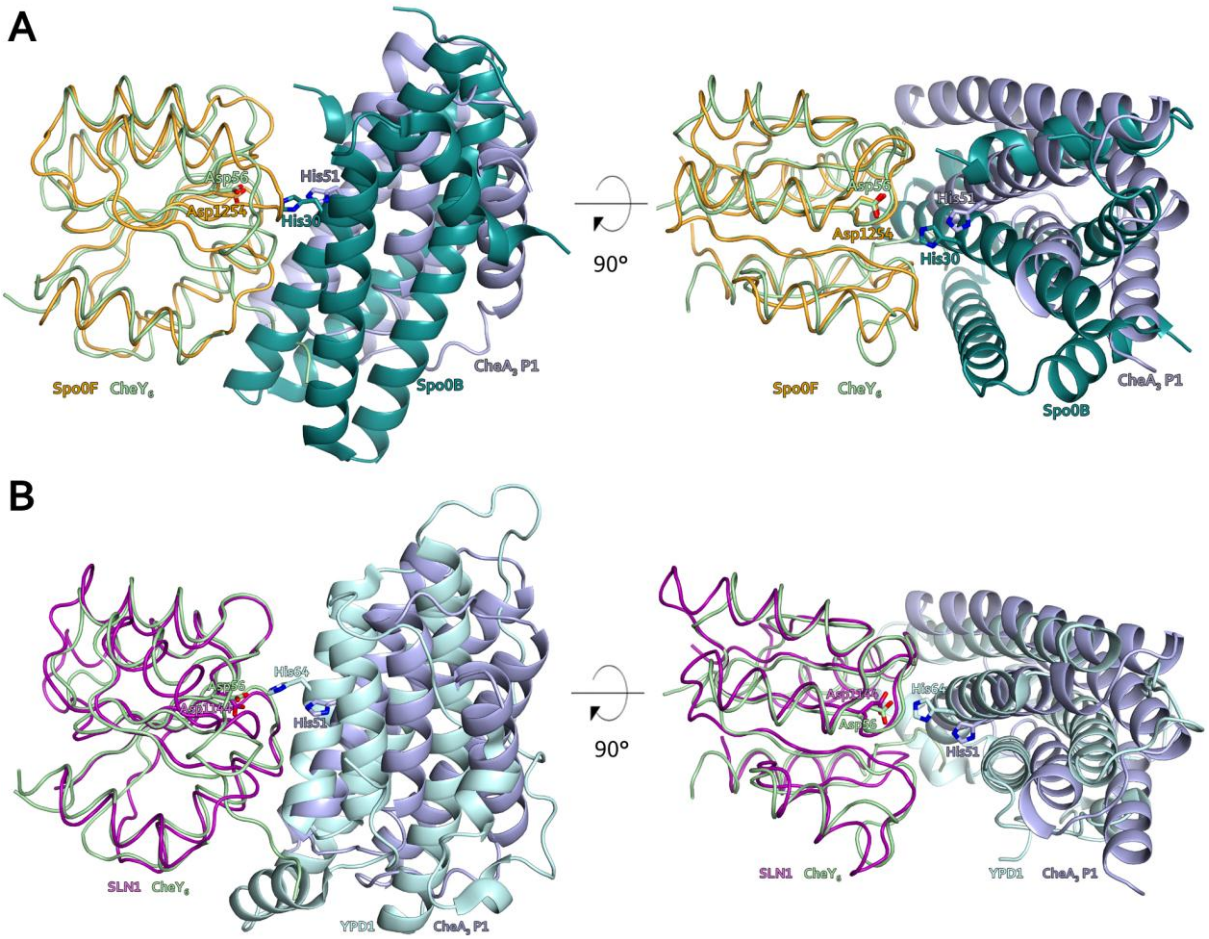


Figure 3.13: Superposition of the $CheA_3P1$-$CheY_6$ complex structure with the Spo0B-Spo0F complex and YPD1-SLN1 complex. (A) Superposition of $CheA_3P1$-$CheY_6$ with Spo0B-Spo0F (PDB-code 1F51). (B) Superposition of $CheA_3P1$-$CheY_6$ with YPD1-SLN1 (1OXB). All structures were superimposed on the response regulators using SSM as implemented in COOT [124]. $CheY_6$ is shown in palegreen, $CheA_3P1$ in lightblue, Spo0F in orange, Spo0B in teal, SLN1 in purple and YPD1 in aquamarine. Only the four-helix bundle of the Spo0B dimer is shown, the C-terminal domain is omitted for clarity. The orientation is similar to Fig. 3.3. $CheA_3P1$ is structurally more similar to the monomeric YPD1 than to Spo0B.

$CheA_3P1$-$CheY_6$ shows the smallest binding interface amongst the three complexes solved so far with 530 $Å^2$ compared to 953 $Å^2$ in YPD1/SLN1 and 1200 $Å^2$ in Spo0F/Spo0B. The interface is smaller since $CheA_3P1$ has neither the C-terminal α/β fold domain seen in Spo0B nor the additional N-terminal helix seen in YPD1. Despite this small interface, $CheA_3P1$ shows binding and efficient phosphotransfer to $CheY_6$. It was previously proposed that the P2 domain of CheA-like proteins might be necessary for binding of the response regulator to CheA by increasing the binding interface [159]. However, $CheA_3$ does not have a P2 domain and the experimental data presented here and elsewhere [23] show that the P1 domain alone is sufficient for binding of and specific phosphotransfer to the cognate response regulator. Moreover, the structural and mutational analysis suggests that their interaction is mediated by very distinct residue clusters on the response regulator and the Hpt domain. The former being located on the N-terminal region of α1 and the loop region between β5 and α5 and the latter being located on the N- and C-terminal region of αA and αB, respectively, and the loop region connecting αB and αC. This is in good agreement with a recently published computational analysis of amino-acid coevolution of cognate histidine kinase-response regulator pairs [154] and the binding site found in the YPD1/SLN1 [159,160] or Spo0B/Spo0F [157,158] complex.

The specificity of protein-protein interactions is essential for most cellular processes. Despite the vast number of these interactions our understanding of the molecular basis of their specificity is limited. The structure presented here allowed to identify specificity determinants for the CheA-CheY interaction. This information was then successfully used to redesign non-cognate response regulators to allow them to be rapidly phosphorylated by

$CheA_3P1$-P (Figure 3.11 and 3.12). While the Laub group have reengineered a HPK to specifically phosphorylate non-cognate RR substrates [154], the work presented here now shows that it is possible to rationally reengineer a RR so that it can be phosphorylated by a non-cognate HPK. The rewiring of the phosphotransfer specificity described here represents the first example of the redesign of the intracellular part of the chemotaxis pathway and provides valuable insight into how cells mediate specificity in one of the most abundant signal transduction pathways, two-component signalling.

The ability to reengineer phosphotransfer specificity coupled with work from the Bourret group [161], which has shown that response regulator autodephosphorylation rate can be manipulated, provide a platform for the future design of synthetic two-component circuits with customizable kinetics.

Chapter 4

Crystal structures of $CheY_3$ - a new paradigm for response regulator activation?

Chapter 4 Crystal structures of $CheY_3$ - a new paradigm for response regulator activation?

4.1 Summary

Response regulators play important roles in a plethora of signalling pathways regulating e.g. transcription, two-component signalling and cyclic di-GMP pathways. Most of them include an output domain, e.g. a DNA binding motif, and a regulatory receiver domain that acts as phosphorylation-dependant molecular switch. Structurally, all receiver domains share a conserved $(\alpha/\beta)_5$ topology. Upon phosphorylation of a conserved Asp residue in the active site, there is a conformational change that alters the α4-β5-α5 face of the receiver domain and allows signal transduction. The mechanism of this conformational change has been described in detail and has been shown to be crucially dependant on a conserved Thr-Tyr coupling adjacent to the active site. Here, two crystal structures of the chemotaxis response regulator $CheY_3$ from R. sphaeroides are presented. The two structures have distinct conformations in the α4-β5-α5 area and likely represent inactive and active conformation of $CheY_3$. Biophysical and cellular analyses of structure-guided mutations are presented and a model for activation that does not include Thr-Tyr coupling but rather reorientation of α4 is proposed.

4.2 Introduction

Two-component signalling systems affect a wide variety of cellular processes primarily through regulation of gene expression but also through protein-protein interactions. An essential part of these systems are response regulators. All response regulators share a common regulatory receiver domain but differ in their associated output domain [162]. A great majority of these output domains contain DNA-binding motifs often featuring helix-turn-helix motifs [163] as exemplified by the structures of members of the OmpR/PhoB and NarL/FixJ families [164,165]. However, a significant fraction of output domains has enzymatic functions and can play crucial roles in adaptation (e.g. the methyltransferase CheB [27]) or cyclic di-GMP signalling (e.g. diguanylate cyclases [166]). In addition, output domains can also act through protein-protein interactions [167]. Response regulators can also consist of a receiver domain alone without an associated output domain. These so called single-domain response regulators can for example be found in chemotaxis, regulating bacterial motility. In fact, the best studied response regulator, CheY from *E. coli,* comprises of just the receiver domain and is essential to alter the swimming direction of the bacterium as a response to extracellular stimuli.

The receiver domain of response regulators is conserved in structure and function. It adopts a typical $(\alpha/\beta)_5$ fold with a central five-stranded beta-sheet surrounded by five helices [168]. Receiver domains act as molecular switches that are triggered by phosphorylation of a conserved Asp residue. In addition to this Asp, the active site is highly conserved and includes a Mg^{2+} binding site in which the metal ion is coordinated by a carbonyl oxygen and two Asp residues, one of which is the phosphoryl acceptor. In the active conformation the

divalent cation stabilizes the receiver domain by compensating for the additional negative charge of the phosphate group. Phosphorylation ultimately leads to a conformational change in the α_4-β_5-α_5 face of the response regulator which is often exploited to control output function through interaction with other domains or proteins [169]. In *E. coli* CheY this area of the protein is responsible for interaction with its effector protein, the flagellar motor component FliM [170]. As a result of the conformational change the affinity of active, phosphorylated CheY to FliM is approximately 20-times tighter than that of inactive CheY [171]. The mechanism of response regulator activation has been subject of intense study since the first structural characterization of an active receiver domain [172]. There are several key differences between the active and inactive form of the response regulator. Upon phosphorylation, a conserved Thr or Ser residue forms a hydrogen bond with a phosphoryl group oxygen, thus repositioning its sidechain and with it the $\beta_4\alpha_4$ loop. This allows a conserved aromatic residue (usually Tyr) on β_5 to adopt a different rotamer, thereby burying its side chain in a hydrophobic pocket between β_4 and α_4 [168]. In the inactive, open conformation the aromatic residue is facing the solvent and when superposed with the FliM-complex structure would sterically clash with the interacting N-terminal FliM helix. Molecular dynamics simulations in conjunction with spectroscopic analysis using NMR suggest that the mechanism of this Thr-Tyr coupling is mostly mediated by changes in α_4. These include a tilt, rotation, and register shift accompanied with a loss and gain of one half helical turn on opposite ends of this helix [173,174].

The aim of this project was to characterize the single-domain response regulator $CheY_3$ from *R. sphaeroides* that is essential for chemotaxis. The crystallographic analysis revealed two

conformations that likely resemble active and inactive conformation of the receiver domain. To verify the structural insights a series of biophysical and cellular experiments were carried out leading to the proposition of a novel mode for response regulator activation.

4.3 Expression and purification of $CheY_3$

A construct containing the full length response regulator $CheY_3$ (residues 1-121) from *R. sphaeroides* cloned into the bacterial expression vector pQE60 was obtained from Dr. Steven L. Porter. $CheY_3$ was expressed in M15 cells following the protocol described in Section 2.2.1. Initially the purification was performed as discussed in section 2.2.1 using Ni^{2+}-affinity chromatography as a first purification step followed by gelfiltration. Despite a high purity of >95 % as judged from SDS-PAGE the yield of purified protein was poor mostly due to a lot of precipitation at the concentration step following the Ni^{2+}-affinity purification. The role of buffer composition used in the affinity and size-exclusion chromatography steps was evaluated as a possible cause of the precipitation of $CheY_3$. A ThermoFluor screen was carried out (see Section 2.10) in which the protein is mixed with 48 different buffer and salt combinations plus 48 different additives such as nucleotides or divalent cations together with a fluorescent dye that binds to strongly exposed hydrophobic patches of a protein. These occur mostly upon unfolding, thereby exposing the hydrophobic core. Upon binding to the protein the fluorescence spectrum of the dye changes which therefore creates a folding-specific probe. The experiment is carried out in a Real-Time PCR machine that monitors the change of fluorescence of this dye in reference to the temperature. Surprisingly, buffers without salt and at a low pH dramatically increase the thermostability

of the protein. The most positive effect could be found for Sodium-Acetate (NaAc) at pH 4.0 (Fig. 4.1a). The melting temperature compared to the initial purification buffer was increased by 8 °C. In addition, changing the divalent cation from Mg^{2+} to Ca^{2+} or Mn^{2+} improved the thermostability by approximately 10 °C (Fig. 4.1b). The purification was subsequently carried out with the initial buffer for the Ni^{2+}-affinity column followed by rapid dilution into 20 mM NaAc, pH 4.0, 2 mM $CaCl_2$. The final gelfiltration step was then carried out in 20 mM NaAc, pH 4.0, 2 mM $MnCl_2$, since Mn^{2+} is more suitable for crystallization. The yield could be improved by one order of magnitude and the protein was stable for > 1 year at room temperature. The purified protein was also reasonably stable in a neutral pH buffer upon addition of 5 mM NaAc and 2 mM $MnCl_2$.

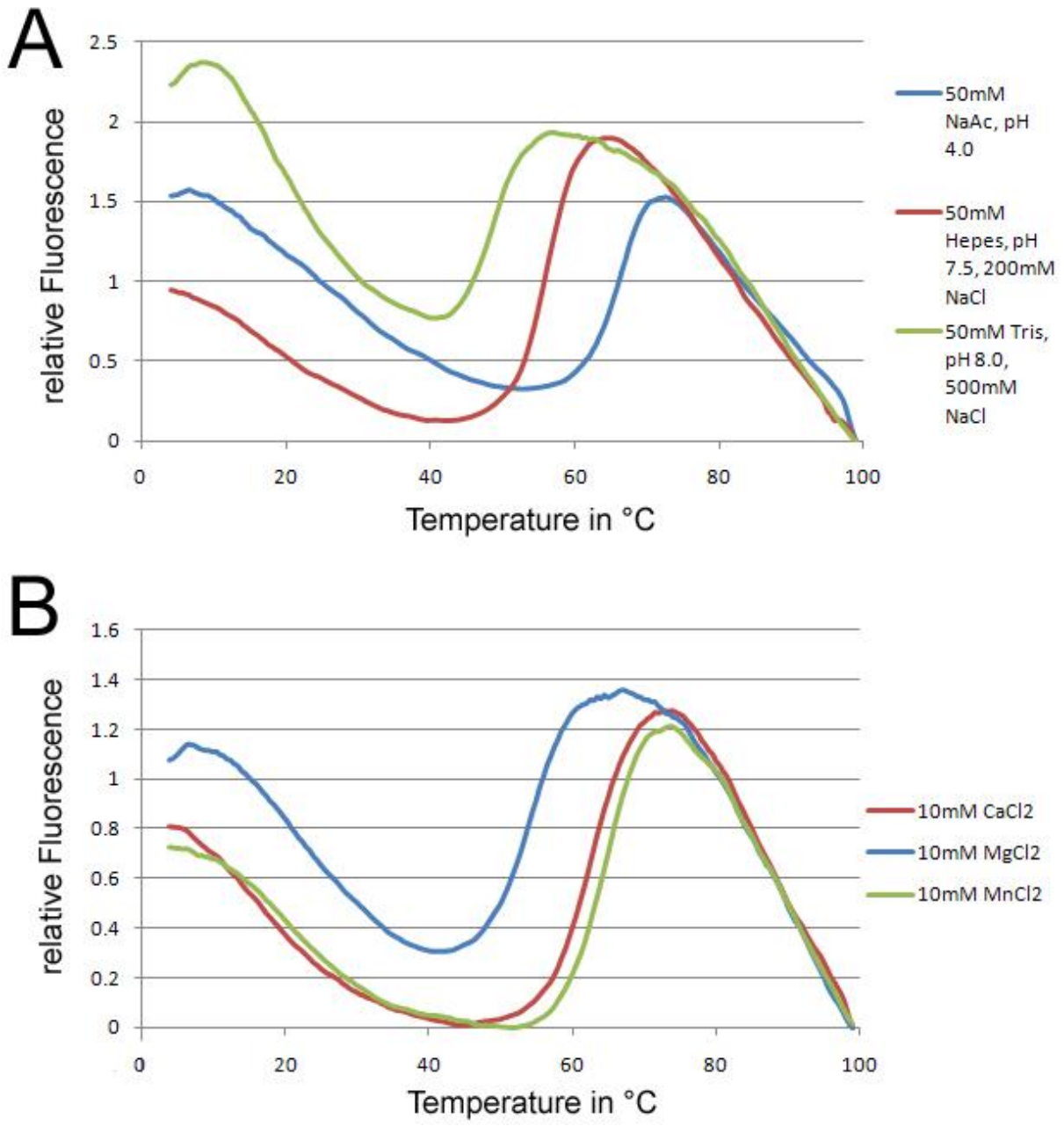


Figure 4.1: ThermoFluor assay for $CheY_3$. Selected buffers are shown in (A), selected additives in (B).

4.4 Crystallization and structure determination

$CheY_3$ crystallized in 20 % PEG 3350 and 200 mM potassium acetate. This condition was optimized using a dilution series of the original buffer composition and by use of an additive

screen. Optimized crystals reached their final size of 1000x200x200 μm^3 after 48 hrs (Fig. 4.2). Diffraction data were collected at 100 K with the crystals being flash-cooled in a cryo N_2 gas stream. Before flash-freezing, crystals were cryo-protected with perfluoropolyether oil PFOX125/03 (Lancaster Synthesis). In addition to the native $CheY_3$ protein, crystal trials were also set-up with a $CheY_3$ sample to which $BeCl_2$ and NaF were added. The resulting BeF_3^- complex associates with the active site Asp of the response regulator in a stable manner and mimics the phosphoryl group and thus the active state of the receiver domain [175]. Phosphorylated response regulators have very short half-lives due to their auto-dephosphorylation activity [168] and are therefore not suitable for crystallization. The BeF_3^--bound response regulators are stable in solution and are can thus be used to characterize the active conformation crystallographically. The BeF_3^--modified $CheY_3$ crystallized in the same condition as the native protein and was optimized, frozen and cryo-protected in the same way.

Two data sets to a resolution of 1.20 Å (native) and 1.30 Å (BeF_3^--modified) were collected at beamlines I02 and I03, respectively, at Diamond Light Source (Fig. 4.2). Due to the high resolution, several datasets with different high-resolution diffraction limits and exposure times were collected. The crystals belong to space group $P2_12_12_1$ with a solvent content of 42 % (native) and 34 % (BeF_3^--modifed).

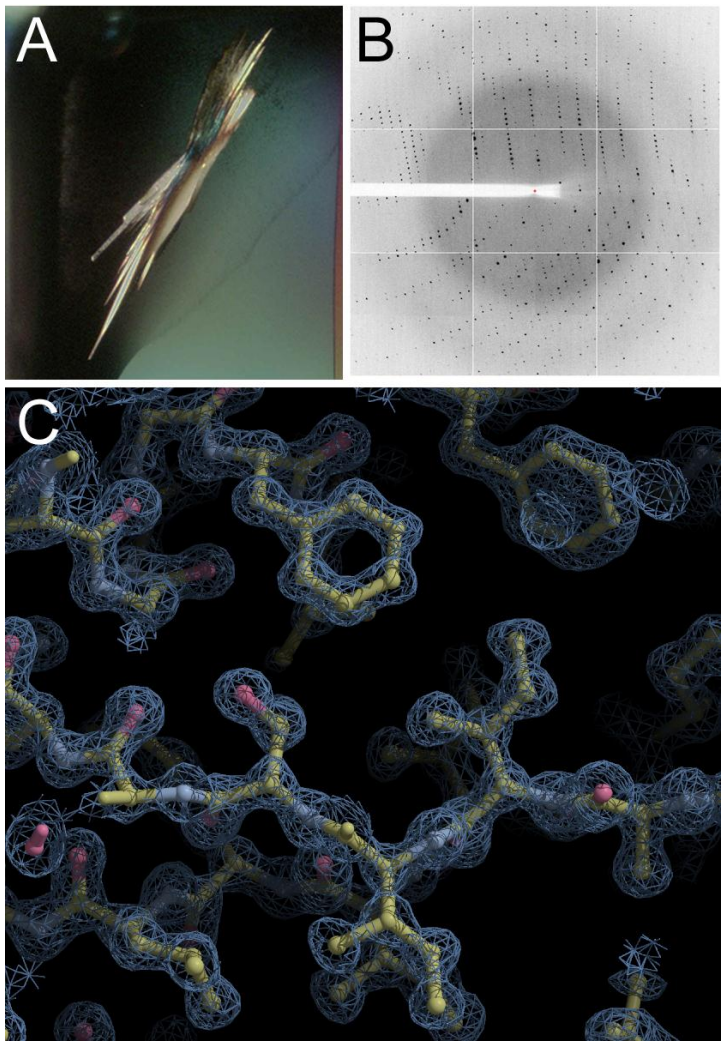


Figure 4.2: Data collection and structure solution of $CheY_3$. A) Optimized crystal of native $CheY_3$. B) Representative diffraction image for the native, high-resolution dataset. C) Representative $2F_{obs}$-F_{calc} map contoured at 1.5σ of the final, refined model.

Initial attempts to solve the structure by molecular replacement using Phaser [122] failed. The reason for this is probably the low sequence identity of < 30 % between $CheY_3$ and the various response regulator models that were used. In addition to Phaser, an automated approach for molecular replacement using the CaspR server [176] was also tried. CaspR uses a

combination of well-established stand-alone software tools in order to overcome low sequence similarity. The program "begins with the generation of multiple structure-sequence alignments produced with T-Coffee [177], followed by homology model building using MODELLER [178], molecular replacement with AMoRe [179], and model refinement based on CNS" [180]. Several different response regulator structures were used as input, and a modified version of the crystal structure of the receiver domain of PhoP from *B. subtilis* [181] yielded a successful molecular replacement solution. Iterative positional refinement was carried out in autoBUSTER [125] and Phenix [126] with manual model building and adjustment in COOT [124].

Stereochemical properties were assessed by MOLPROBITY [127]. Ramachandran statistics are as follows (favored/disallowed (%)): $CheY_3$ native: 97.5/0.0, $CheY_3$ BeF_3^--modified 99.2/0.0. Data collection and refinement statistics can be found in Table 2.

Table 2. Data collection and refinement statistics

	$CheY_3$ native	$CheY_3$ BeF_3^--modified
Data collection		
Resolution (Å)	1.20 (1.24 - 1.20)	1.30 (1.35 - 1.30)
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimension		
a, b, c (Å)	34.6, 42.4, 69.4	42.6, 42.4, 66.0
α, β, γ (°)	90, 90, 90	90, 90, 90
Redundancy	5.5 (4.6)	7.2 (4.3)
Completeness (%)	95.0 (85.6)	98.9 (91.2)
R_{merge} (%)	9.7 (36.1)	6.3 (65.6)
Avg I/σ	9.8 (3.6)	29.2 (1.9)
Refinement		
Resolution	31 - 1.20 (1.23 - 1.20)	36 - 1.3 (1.34 - 1.30)
No. Reflections	31,107	29,708
R_{work}/R_{free}	16.4 (23.4)/19.8 (30.2)	15.8 (27.0)/19.0 (34.1)
No. Atoms		
Protein	954	935
Sugar	125	185
B-factors		
Protein	26.8	23.3
Water	42.4	36.3
r.m.s. Deviations		
Bond lengths (Å)	0.011	0.010
Bond angles (°)	1.040	1.090

[a] Numbers in parenthesis are for the highest resolution shell

[b] R_{free} equals the R_{work} against 5% of the data removed prior to refinement.

4.5 Crystal structure of $CheY_3$

$CheY_3$ crystallized in two different unit cells both belonging to the orthorhombic spacegroup $P2_12_12_1$. The overall structure of $CheY_3$ in both crystal forms resembles the typical response

regulator topology with a central five stranded β-sheet surrounded by five α-helices (Fig. 4.3 and Fig. 4.4) [182]. Within the active site the phosphorylatable Asp53 in the centre together with Asp10 and the backbone carbonyl of Asn55 coordinate the divalent cation, which in the crystal structure is Mn^{2+}. This architecture has been observed in all response regulator structures so far and represents a common characteristic of these proteins.

Surprisingly, the BeF_3^--modified crystal form did not show any additional density in the active site. The most likely reason for this is that the BeF_3^- modification could not be accommodated in the crystal lattice and thus only the non-modified, native $CheY_3$ crystallized. Since the BeF_3^- modification is transient and non-covalent, the equilibrium between BeF_3^- bound and unbound $CheY_3$ can be shifted towards the latter upon crystal formation. In support of this hypothesis, there is a crystal contact between two symmetry related molecules that involves the active site and indeed would sterically clash with the proposed BeF_3^- binding site. This crystal contact is found in both crystal forms. The side chain of Glu24 inserts into the active site of a symmetry-related $CheY_3$ copy, coordinating the divalent cation on one side and forming a hydrogen bond with Thr84 on the other side (Fig. 4.5). As a result the base of the loop region connecting β4 and α4 is locked in a discrete conformation. In order to exclude a possible effect of this contact on the protein conformation, Glu24 was mutated to Ser. The mutant protein was solubly expressed, however extensive crystallization trials did not yield any crystals. Therefore, the crystal contact mediated by Glu24 is essential for crystal formation and possible effects on the protein conformation cannot be excluded.

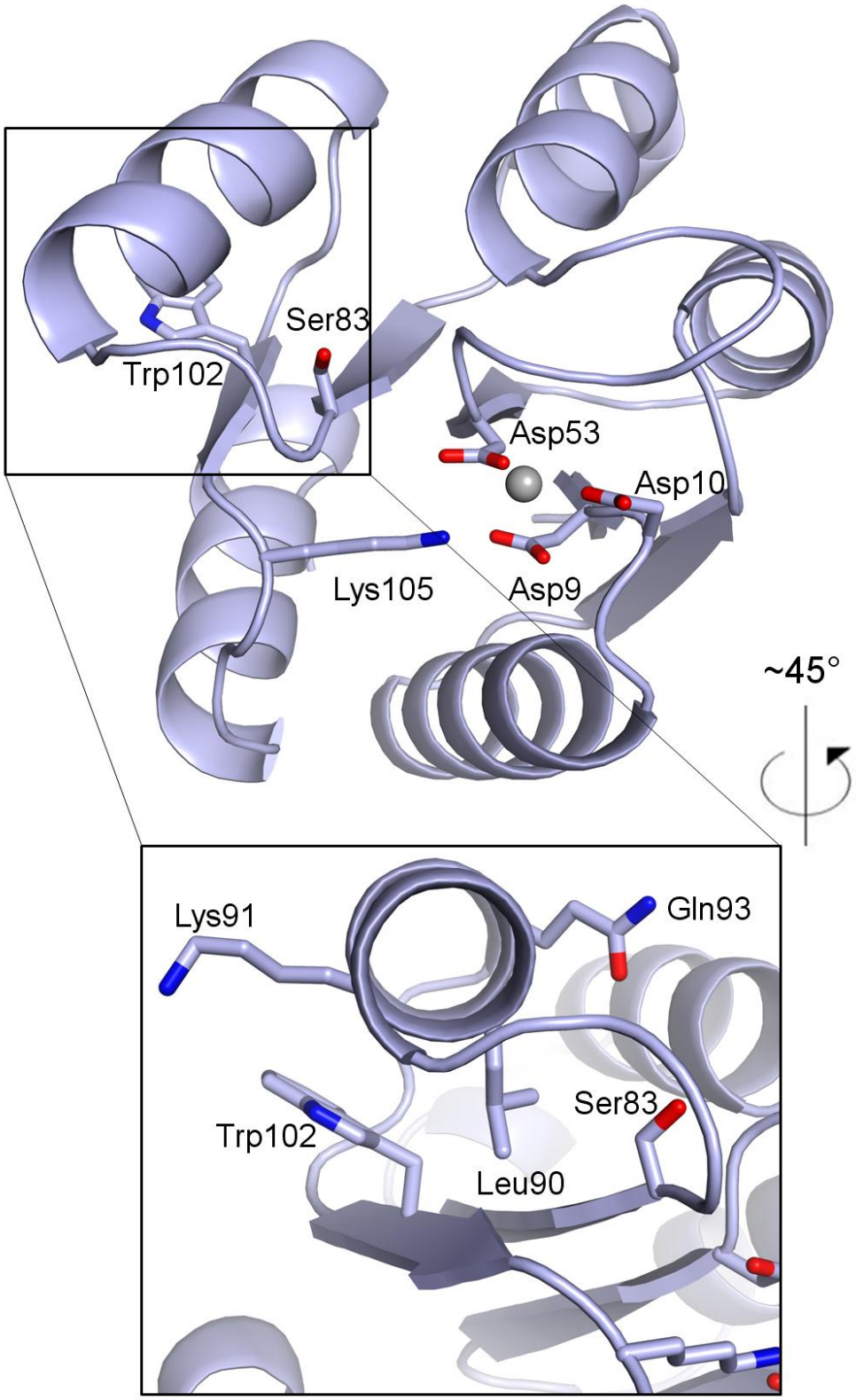


Figure 4.3: Crystal structure of unmodified, native $CheY_3$. Residues involved in the active site and the activation mechanism are shown as sticks. Mn^{2+} is shown as grey sphere.

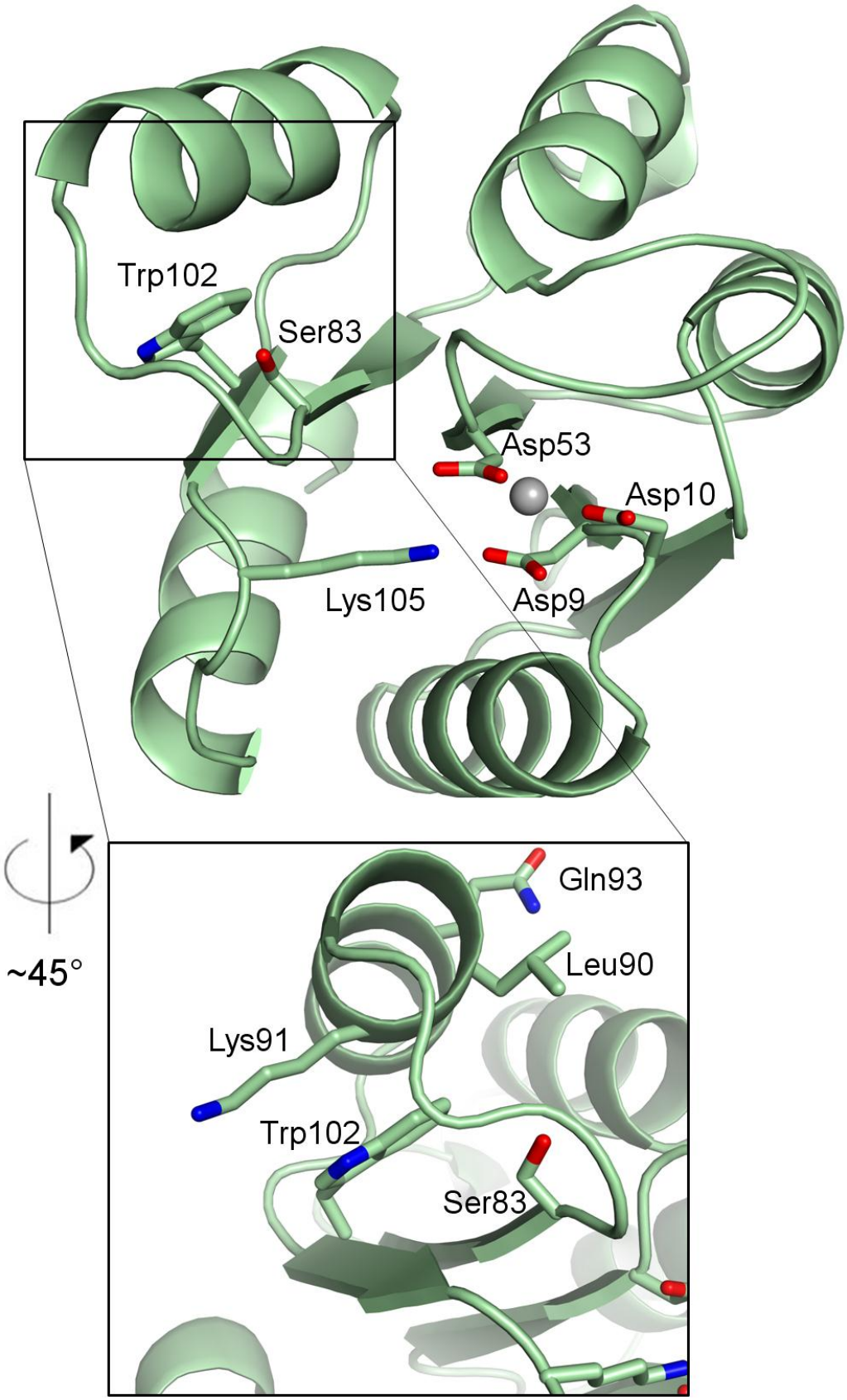


Figure 4.4: Crystal structure of BeF_3^--modified $CheY_3$. Residues involved in the active site and the activation mechanism are shown as sticks. Mn^{2+} is shown as grey sphere.

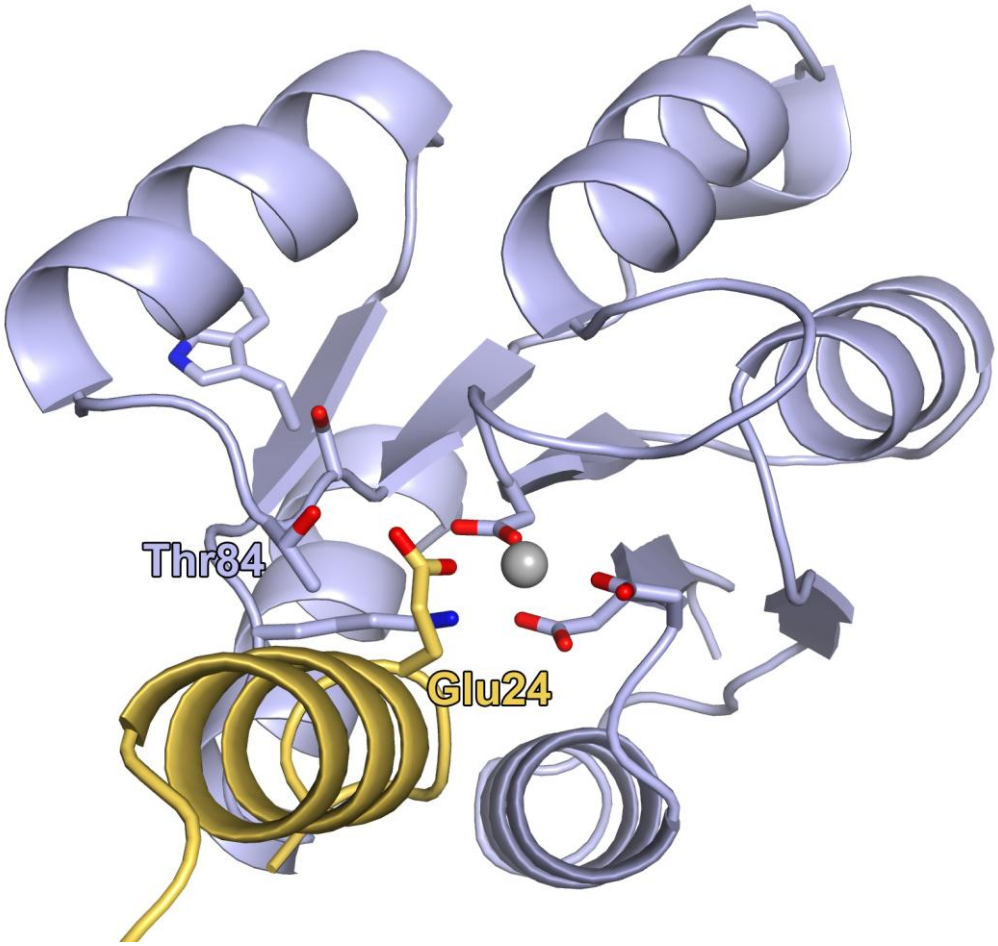


Figure 4.5: An essential crystal contact blocks the active site. The native $CheY_3$ structure is shown in lightblue and a symmetry-related copy in yelloworange. The crystal contact is mediated by Glu24, Thr84 shown in stick representation and the divalent cation shown as grey sphere.

Despite having similar overall structures the two crystal forms show remarkable differences in the conformation of α-helix 4 and the residues on β-strand 5 (Fig. 4.3 and Fig. 4.4). In the native crystal structure Ser83 is oriented towards the active site (Fig. 4.3). Leu90 occupies a hydrophobic pocket formed between β4 and β5 and Gln93 forms a hydrogen bond with Ser83 mediated by a water molecule. Most prominently, Trp102 points out to solvent and adopts an exposed conformation.

In the BeF_3^--modified structure Ser83 points away from the active site. Trp102 is buried in the hydrophobic cavity that was occupied by Leu90 in the native structure. In addition, its aromatic side chain stacks against the aliphatic groups of Lys91 therefore further shielding Trp102 from solvent and stabilizing its conformation. The hydrogen bond between Gln93 and Ser83 is lost. Due to the characteristic position of the Trp102 side chain, the native structure is hereafter called 'exposed' (Fig. 4.3) and the BeF_3^--modified structure is referred to as 'buried' (Fig. 4.4).

The different conformation of the above mentioned residues is accompanied by elongation, tilt and rotation of α4 (Fig. 4.6). In the exposed structure, Ser86 is the first residue to take part in the typical α-helix hydrogen bond pattern forming a hydrogen bond between the Ser86 carbonyl oxygen and the Leu90 amide nitrogen. In the buried structure, the first residue to do so is Ala87 (Ala87 carbonyl oxygen to Lys91 amide nitrogen). Thus, α4 in the exposed structure is elongated by one residue or almost half a helical turn. This elongation also tilts α4 so that it is closer to β-strands four and five. To accommodate the additional hydrogen bond at the N-terminus of α4, the helix rotates by about 30°, thereby allowing Leu90 to occupy the hydrophobic pocket formed between β4 and β5. In summary, the reorientation of α4 significantly alters the surface of this helix by exposing different residues and also has a dramatic effect on the conformation of Trp102 (Fig. 4.6).

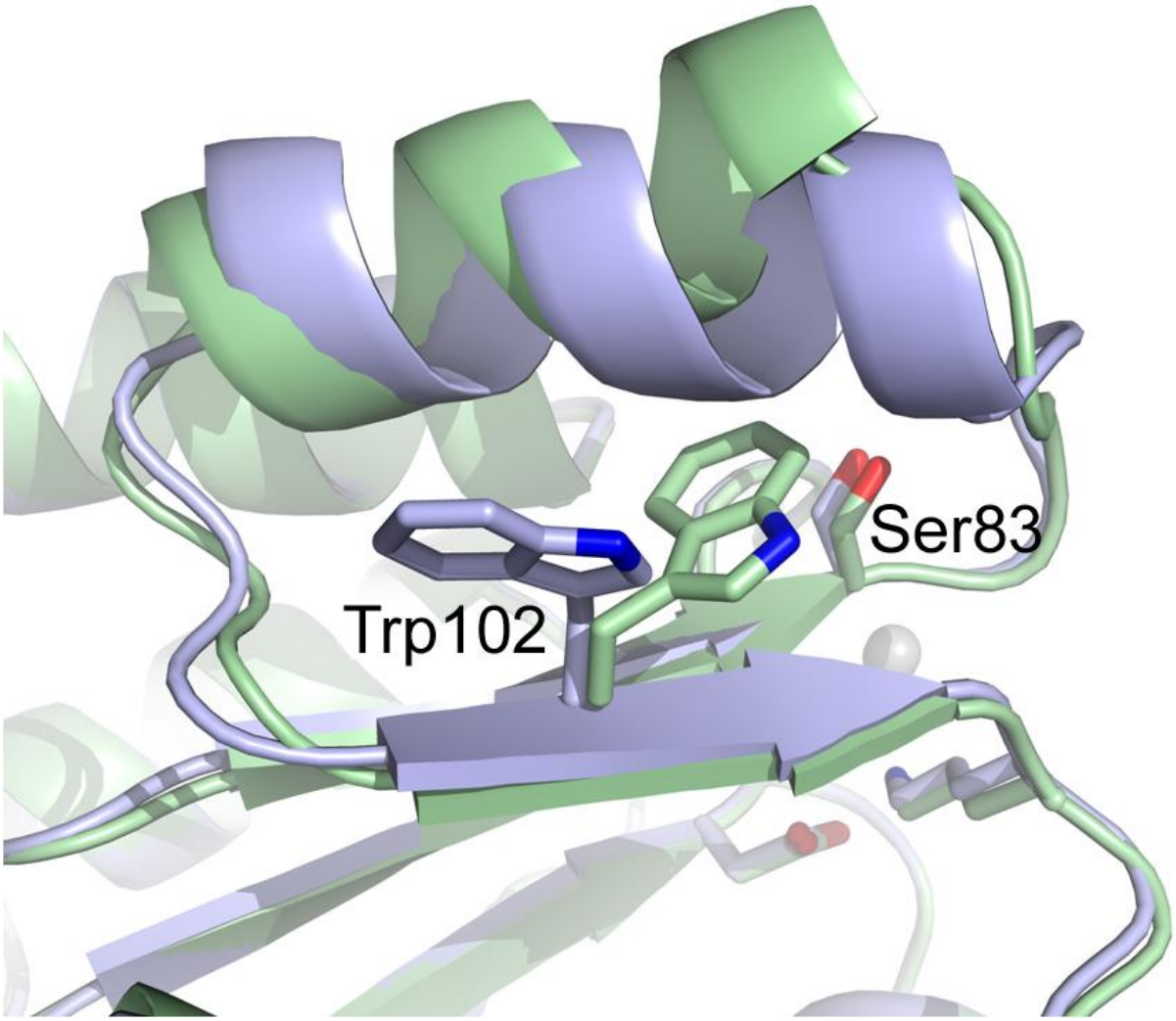


Figure 4.6: Detailed view onto α4 and β5 in the exposed and buried conformation. The structures were superposed using SSM as implemented in COOT [124]. Colour coding follows Fig. 4.3 and 4.4.

4.6 Biophysical and functional characterization of $CheY_3$

Previous studies reported that the fluorescence of the only Trp in $CheY_3$, Trp102, is dependent on the activation state of the response regulator [26]. Indeed, repeating this assay showed that upon addition of the small molecule phosphodonor acetyl-phosphate, the fluorescence is quenched suggesting that the environment of Trp102 changes (Fig. 4.7). A likely reason for fluorescence quenching is a reorientation of the aromatic residue from a

buried to a solvent-exposed conformation. Due to collision with solvent molecules the number of excited aromatic π-systems decreases and with it the fluorescence. Interestingly, the two crystal structures presented here show Trp102 in exactly those conformations, one is solvent–exposed (Fig. 4.3) and the other is buried and shielded from solvent (Fig. 4.4). It was therefore hypothesized whether the two crystal structures represent the active and inactive conformation, respectively.

In order to investigate this possibility further, several mutants were generated and tested in the Trp102 fluorescence assay (Fig. 4.7). Ser83 has been shown to be crucial for response regulator activation since it interacts with the phosphoryl group and initiates the conformational changes on the α4-β5-α5 face. Mutation of this residue to Asp abolished a change in fluorescence completely, whereas mutating this residue to Ala only had an effect on the kinetics of the fluorescence quenching. This suggests that Ser83, and especially its interaction with the phosphoryl-group, is important but not crucial for the reorientation of Trp102. Both amplitude and rate of the fluorescence quenching cannot be linked to function or structure since there are too many factors (e.g. solvent accessibility, change in hydrophobicity of the Trp environment, small local conformational changes) that cannot be monitored and affect these variables. In addition, the effect of the length of α4 was investigated by introducing Pro at positions 90 and 91, the first residues to form α-helical amide hydrogen bonds in the respective structures. Pro is a helix-breaker since its amide nitrogen cannot form hydrogen bonds. Thus, introduction of this residue restricts α4 to end at positions 90 and 91, respectively. The Leu90Pro showed the same quenching effect as the wild-type protein, whereas the Lys91Pro mutation did not show any change in fluorescence

upon addition of acetyl-phosphate. The latter effect could be either a result of the shortening of the helix and with it the ability of Trp102 to change its orientation or due to lack of the solvent-shielding effect of the Lys side chain (Fig. 4.4).

The Trp fluorescence assay can monitor the orientation of the Trp102 side chain and with it the activation state of the wild-type protein. However, the function of $CheY_3$ which is mediated by its interaction with FliM via the α4-β4-α5 face of the response regulator is not solely dependent on the conformation of the aromatic residue on β5 but rather the entire α4-β4-α5 surface. Consequently, the results of the Trp fluorescence assay for the investigated mutants had to be validated in an independent, functional assay to correlate the effect of the mutation on the Trp orientation with its effect on the entire α4-β4-α5 surface.

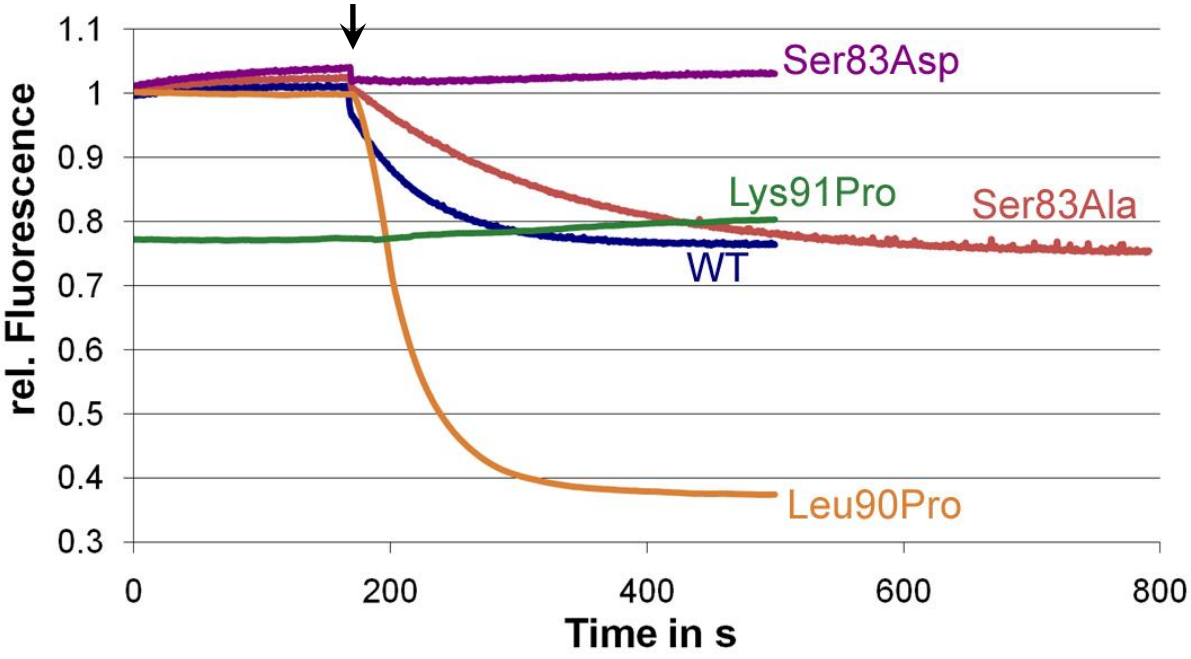


Figure 4.7: Trp fluorescence assay. The fluorescence of Trp102 is plotted against time. The arrow marks the activation of the response regulator by addition of acetyl-phosphate.

The function of $CheY_3$ can be monitored in a swarm plate assay (Fig. 4.8). In this assay the chemotactic response and the motility of bacteria can be studied. Here, a $CheY_3$-$CheY_4$ deletion strain was used since these proteins are redundant for each other. In this background, mutant and wild-type $CheY_3$ constructs were then expressed from an IPTG-inducible plasmid. This set-up allowed to directly monitor the effect of a mutation in $CheY_3$ on its function as expressed in the chemotactic behaviour of *R. sphaeroides* on the swarm plate. As expected, expression of wild-type $CheY_3$ can restore the chemotactic response of *R. sphaeroides* and rescue the $CheY_3$-$CheY_4$ deletion (Fig. 4.8). All of the mutants studied in the Trp fluorescence assay however were non-chemotactic (Fig. 4.8). Mutations on α4 or of the crucial switch residue Ser83 lead to a loss-of-function phenotype. This suggests that the $CheY_3$ mutants Ser83Ala and Leu90Pro do not have an effect on the change of orientation of Trp102 upon activation but alter the α4-β4-α5 surface compared to wild-type $CheY_3$. In addition, it also suggests that the conformation of α4 needs to be tightly regulated for response regulator function since restricting the helix by introduction of Pro lead to loss of function.

Since the results of the Trp fluorescence assay and the swarm plate assay differ remarkably, the fluorescence of Trp102 alone cannot be used to elucidate the conformational changes of single residues that occur upon response regulator activation. In order to confirm the hypothesis that the two crystal structures represent active and inactive conformation of $CheY_3$ and that the conformational changes seen in those structures correspond to the changes occurring in solution, $CheY_3$ was studied using NMR. NMR is the ideal tool to study dynamic processes on a single residue level [183]. In particular, since X-ray structures were

available, changes in the chemical shift of a residue can immediately be mapped onto the structure. The experimental set-up as well as the data collection and analysis were performed by Dr. Christina Redfield. The spectra for native and BeF_3^--modified $CheY_3$ were well-dispersed and could be completely assigned (Fig. 4.9). Comparison of the HSQC spectra of the different activation states confirmed that the majority of conformational changes map onto the α4-β4 surface as observed in the two crystal structures. However, further analysis is needed to confirm that the two crystal structures resemble the active and inactive $CheY_3$ structure in solution.

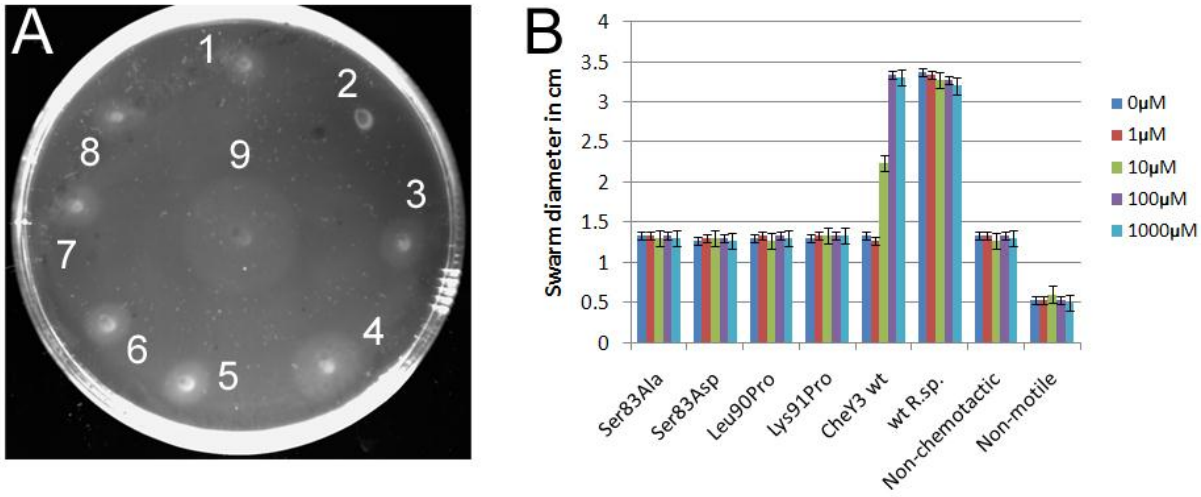
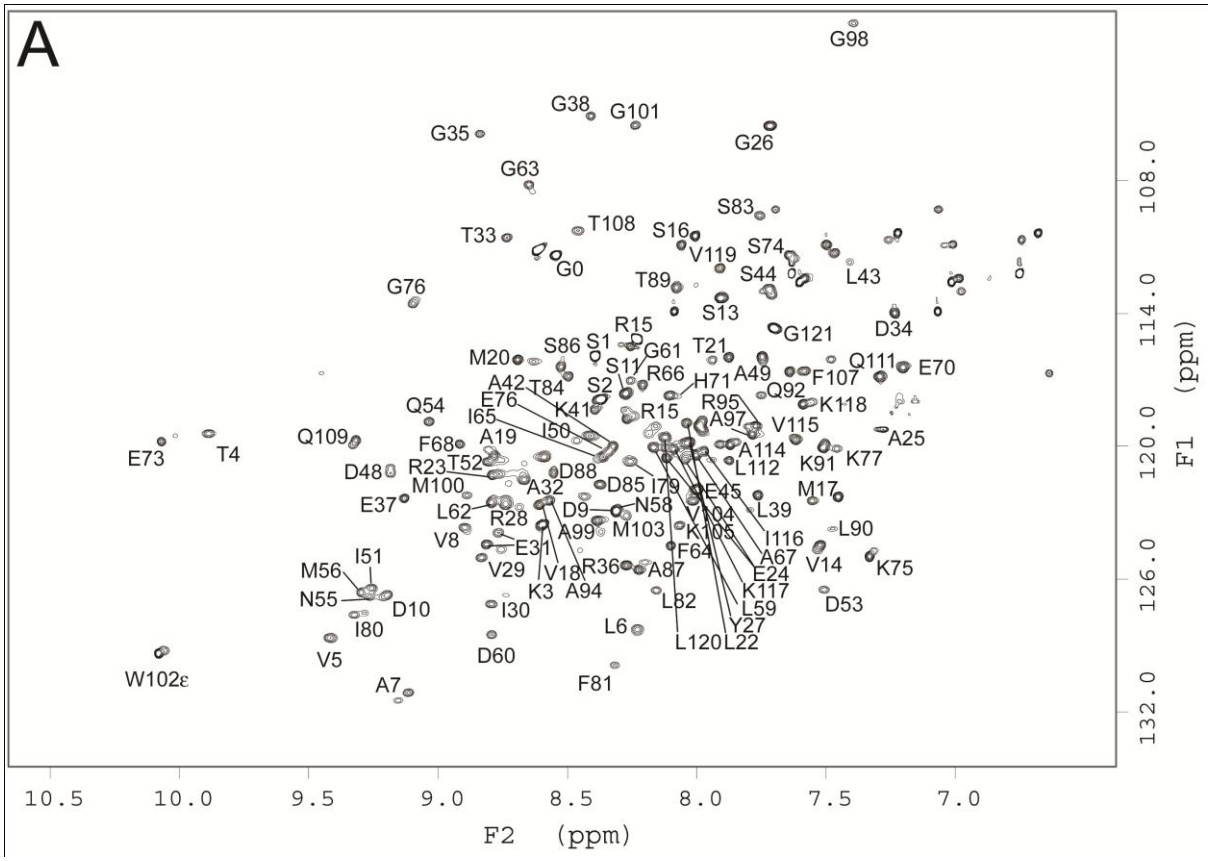


Figure 4.8: Swarm plate assay. A) A representative swarm plate at 10 µM IPTG. 1: Non-chemotactic control; 2: Non-motile control; 3: Control plasmid; 4: wild-type $CheY_3$; 5: $CheY_3$ Ser83Ala; 6: $CheY_3$ Ser83Asp; 7: $CheY_3$ Leu90Pro; 8: Lys91Pro; 9: wild-type *R. sphaeroides*. B) Evaluation of the swarm diameter. Each experiment was done in triplicate. Different IPTG concentrations used for induction are shown in separate colours.


A
G98
G38
G101
G35
G26
G63
S83
T33
T108
S16
V119
S74
G0
T89
S44
L43
G76
S13
R15
G121
D34
S86
S1
G61
T21
M20
S11
R66
A49
Q111
E70
A42
T84
S2
H71
F107
E76
Q92
Q54
K41
R95
K118
I65
R15
A97
V115
Q109
I50
A25
E73
T4
F68
A19
A114
D48
R23
T52
D88
L112
K91
K77
M100
A32
D85
I79
E45
M17
E37
L62
D9
N58
V104
L39
R28
A99
M103
K105
L90
V8
E31
I116
F64
A67
M56
I51
V29
V18
R36
A87
E24
V14
K75
N55
D10
K3
A94
K117
I30
L82
D53
I80
L6
L59
Y27
L120
L22
V5
D60
W102ε
A7
F81
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
F2 (ppm)
132.0
126.0
120.0
114.0
108.0
F1 (ppm)

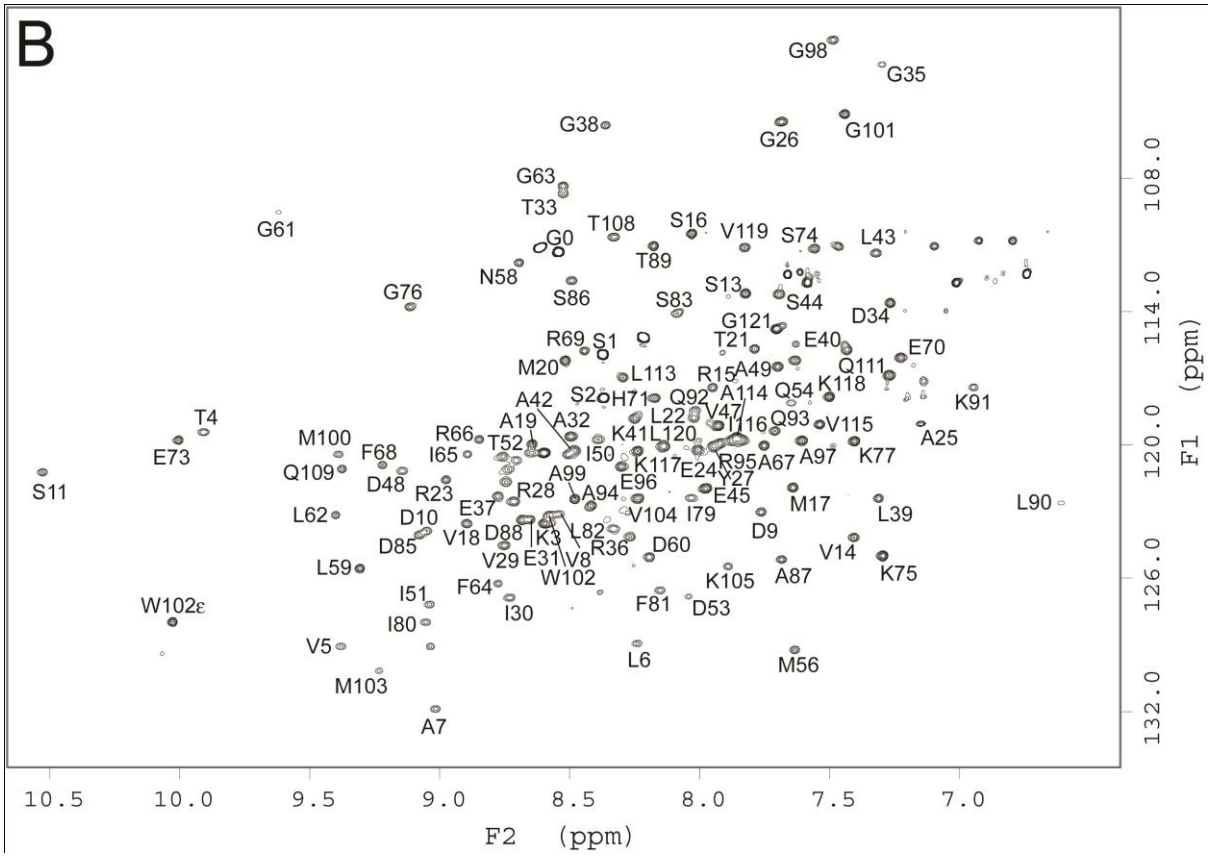

B
G98
G35
G38
G26
G101
G63
T33
G61
T108
S16
V119
S74
L43
G0
T89
N58
S13
G76
S86
S83
S44
D34
G121
R69
S1
T21
E40
E70
M20
L113
R15
A49
Q111
A42
S2
H71
Q92
A114
Q54
K118
K91
T4
A19
A32
L22
V47
Q93
V115
M100
R66
T52
K41
L120
I116
A25
E73
F68
I65
I50
R95
A67
A97
K77
Q109
A99
K117
E24
S11
D48
R23
R28
E96
Y27
E45
M17
L90
L62
E37
A94
V104
I79
L39
D10
D9
D85
V18
D88
K3
L82
R36
D60
V14
V29
E31
V8
W102
K105
A87
K75
L59
F64
I51
F81
D53
W102ε
I30
I80
V5
L6
M56
M103
A7
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
F2 (ppm)
132.0
126.0
120.0
114.0
108.0
F1 (ppm)

Fig. 4.9: HSQC spectra of ^{15}N-labelled $CheY_3$. (A) Native; (B) BeF_3^--modified. F1 describes the ^{15}N channel, F2 the ^{1}H channel. All residues could be assigned and corresponding peaks are labelled.

4.6 Discussion

The function of response regulators as molecular switches relies crucially on the ability to change their conformation upon phosphorylation. This conformational change alters the interface with downstream effectors or output domains and allows signal transduction. Structurally, this interface is formed by the α4-β5-α5 face of the receiver domain. In the extensively studied *E. coli* CheY single domain response regulator the phosphorylation event in the active site results in reorientation of Thr87 which triggers Tyr106 to adopt a buried conformation in between α4 and β4. This so-called Thr-Tyr coupling has been observed in almost all other structurally characterized response regulators. One exception is $CheY_2$ from *Sinorhizobium meliloti* which does not have an aromatic residue on β5 and instead forms a compact α4-β5-α5 surface through additional hydrophobic interactions and reorientation of α4 [184]. Recently, the Kern group studied the activation transition pathway of NtrC using a combination of molecular modelling and NMR [173,174]. They concluded that the molecular mechanism of activation relies on a reorientation of α4 which includes the rearrangement of Tyr of the previously described Thr-Tyr coupling.

Here, two crystal structures of $CheY_3$ from *R. sphaeroides* are presented. Despite having grown in the same crystal condition, the structures adopt distinct conformations, particularly on α4 and β5. These regions are characteristic for the activation state of the response regulator thus it is possible that the two structures resemble active and inactive conformation although none of the two structures has the phosphoryl analogue BeF_3^- bound

to it. There are several lines of evidence in favour of this proposition. The fluorescence of Trp102 is quenched upon phosphorylation, suggesting that this residues changes from a buried to a solvent-exposed conformation as seen in the two crystal structures. In accordance with this, Ser83 is oriented towards the active site in the solvent-exposed structure similar to the orientation of the homologous Thr residues in active CheY and active NtrC. Most importantly, α4 shows a rotation and tilt movement and is elongated by almost half a turn in the solvent-exposed structure. The crystal structure of active CheY, extensive NMR and molecular modelling analysis of active NtrC and the NMR structure of active $CheY_2$ from *S. meliloti*, all show the same movement of α4 [173-175,184]. These findings strongly suggest that the solvent-exposed structure resembles the active conformation and the buried structure the inactive conformation. Interestingly, the orientation of the aromatic residue on β5 (Trp102 in $CheY_3$, Tyr106 in *E.coli* CheY) is opposite in $CheY_3$ (solvent-exposed in the active, buried in the inactive structure) when compared to *E. coli* CheY or all other structurally characterized response regulators with Thr-Tyr coupling. However, the structural changes on α4 that occur upon activation are similar in $CheY_3$ from *R. sphaeroides*, *E. coli* CheY, NtrC or *S. meliloti* $CheY_2$ [173-175,184]. Thus, rather than the orientation of the aromatic residue on β5, the arrangement of α4 should be seen as universal marker for the activation state of a response regulator. In accordance with this the combination of Trp fluorescence studies together with a functional assay reported in this chapter showed that the direction of the aromatic residue is not a marker for functional activation. Instead alterations of α4 lead to a complete loss of function further underlining its importance in response regulator activation.

The fact that both $CheY_3$ structures crystallized in the same condition and showed similar lattice contacts can be explained by the crystal contact formed between Glu24 and the active site of a symmetry related molecule. This interaction locks Thr84 and the base of the β4-α4-loop in a discrete conformation. This loop plays an important role in the activation pathway of other response regulators [185], thus it is likely that the equilibrium between active and inactive conformation that is usually altered by phosphorylation is shifted due to the aforementioned crystal contact.

To exclude a further effect of this contact on the protein conformation and to validate the hypothesis that the crystal structures resemble the active and inactive state of $CheY_3$ in solution, NMR was employed to study the dynamics of $CheY_3$ activation. Both active and inactive form of the response regulator could be assigned and confirmed structural changes on the α4-β4-α5 surface. However, further analysis of the dynamics in particular of α4 upon addition of BeF_3^- is needed to allow a complete understanding of $CheY_3$ activation.

Chapter 5

A dual binding mode for RhoGTPases

in plexin signalling

Chapter 5 A dual binding mode for RhoGTPases in plexin signalling

5.1 Summary

Plexins are cell surface receptors for the semaphorin family of cell guidance cues. The cytoplasmic region comprises a Ras GTPase-activating protein (GAP) domain and a RhoGTPase binding domain. Concomitant binding of extracellular semaphorin and intracellular RhoGTPase triggers GAP activity and signal transduction. The mechanism of this intricate regulation remains elusive. Here two crystal structures of the human Plexin-B1 cytoplasmic region in complex with a constitutively active RhoGTPase, Rac1, are presented. The structure of truncated Plexin-B1-Rac1 complex provides no mechanism for coupling RhoGTPase and Ras binding sites. On inclusion of the juxtamembrane helix, a trimeric structure of Plexin-B1-Rac1 complexes is stabilised by a second, novel, RhoGTPase binding site adjacent to the Ras site. Site-directed mutagenesis combined with cellular and biophysical assays demonstrate that this new binding site is essential for signalling. The findings are consistent with a model in which extracellular and intracellular plexin clustering events combine into a single signalling output.

5.2 Introduction

Plexins constitute a large family of semaphorin receptors that mediate the repulsive chemotactic response necessary for axon guidance in the developing nervous system. They also play an important role in regulating cell migration, angiogenesis and immune responses

[31,186]. Mutations in plexin receptors have been found in cancers from a variety of tissues [187,188].

There are four classes of Plexins (A, B, C and D) [31]. Their architecture is conserved across the family with a large extracellular region including the ligand binding sema domain, a single transmembrane spanning helix, and an intracellular region that transduces signals to a number of downstream pathways [31,186,189]. Recently, truncated ectodomain structures of plexins from different classes in complex with their cognate semaphorin ligands have been solved [35-37]. They revealed a common architecture in which two plexin monomers bind one semaphorin dimer. This bivalency has been shown to be crucial for the function of the plexin-semaphorin complex [35].

Plexins are transmembrane receptors distinguished by their ability to interact directly with small GTPases of the Ras and Rho family through their intracellular region [49,190]. It consists of two domains, the GTPase activating protein (GAP) domain, first identified by sequence similarity to RasGAP, and the RhoGTPase binding domain (RBD) [48,58,191]. Recent structural studies of the intracellular region of human Plexin-B1 and mouse Plexin-A3 revealed that the GAP domain is an integral structural unit, with the RBD forming a domain insertion into one of the exposed GAP domain loops [43,44]. Importantly, the catalytic machinery remained identical, with catalytic arginines found in same positions in RasGAP and both Plexin-B1 and Plexin–A3 [43,44,65,66].

Within the plexin family, the human Plexin-B1 signalling pathway is the most extensively characterized to date; two members of the Ras superfamily have been identified as targets of the Plexin GAP activity so far, R-Ras and M-Ras [49,50]. Inactivation of R-Ras by Plexin-B1

GAP leads to suppression of integrin activation and cell migration, ultimately leading to repulsive axonal guidance [47,192]. Downregulation of M-Ras leads to reduced dendritic outgrowth and branching [50]. The Plexin-B1 RBD has been shown to bind to the Rho GTPases Rnd1, Rac1 and RhoD exclusively in their active, GTP-bound form [193-195]. Small GTPases of the Rho family are key players in remodelling of the actin cytoskeleton and are involved in a plethora of processes initiated by extracellular stimuli [57,196]. Both Rac1 and Rnd1 are important for the ligand-induced activation of the plexin GAP activity and Rac1 has been found to increase semaphorin binding to Plexin-B1 [49,51,197,198].

Simultaneous binding of semaphorin on the extracellular side and a RhoGTPase on the intracellular side is a prerequisite for plexin GAP activity [197,199]. Bivalent semaphorin binding can be mimicked by extracellular, antibody-induced, clustering of the intracellular domain and activation is observed in the presence, but not in the absence of Rnd1 [49,197,199]. This suggests that semaphorins have a crucial role in bringing together plexin receptors as a step towards activation. Despite a number of structural studies on the plexin RBD and its complex with Rnd1 [44,200,201] it remains unclear how RhoGTPases modulate plexins and how the concomitant binding of ligands on the extracellular and the intracellular side of the receptor is integrated into a single signalling output, inactivation of Ras. To address this question the complex between the intracellular region of Plexin-B1 and a constitutively active form of the RhoGTPase Rac1 was characterized both structurally and functionally.

5.3 Expression and purification of Plexin-B1 and Rac1

Several constructs of the intracellular domain of human Plexin-B1 were designed of which three, Plexin-B1$_{cyto}$, Plexin-B1$_{\Delta1}$ and Plexin-B1$_{\Delta2}$, could be solubly expressed in insect cells (Fig. 5.1, see section 2.2.2). The protein was prone to proteolytic digestion leading to four discrete products as shown by SDS-PAGE (Fig. 5.2).

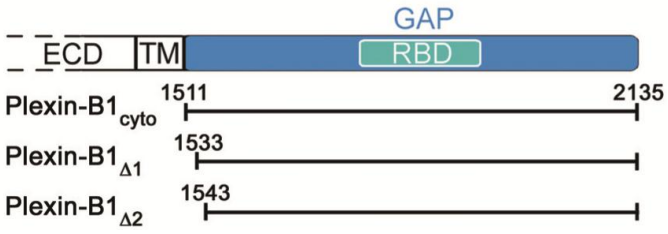


Figure 5.1: Schematic domain organization of the Plexin-B1 receptor. ECD: extracellular domain; TM: transmembrane region; GAP: GTPase activating protein domain; RBD: Rho GTPase binding domain. The constructs used in this study are shown below.

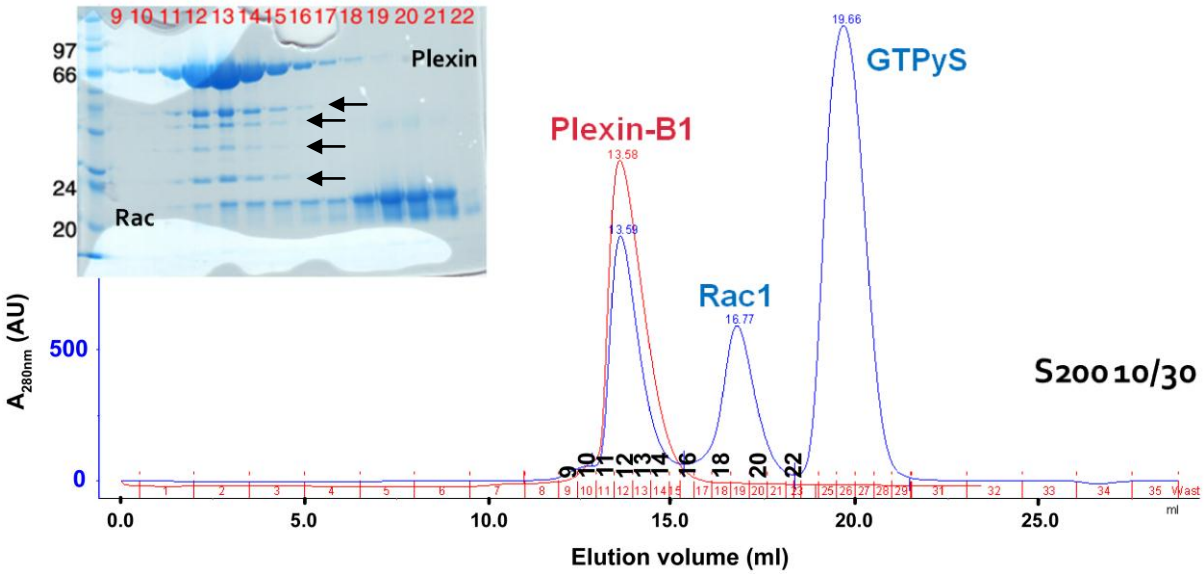


Figure 5.2: Overlay of the gelfiltration elution profiles of Plexin-B1 (red) and a mixture of Plexin-B1, Rac1 Q61L and GTPγS (blue). The inlay shows an SDS-PAGE of the latter run. Fractions are numbered accordingly and the molecular weight of the marker band is given in kDa. The gelfiltration was carried out on a S200 10/30 column. Degradation products of Plexin-B1 are marked by arrows.

These proteolytic fragments eluted from a gelfiltration column at the same time as the full length protein suggesting that the proteolysis did not affect the structural integrity of the construct. To avoid degradation a protease cocktail including AEBSF, leupeptin, pepstatin and E-64 was included in the buffers during lysis and purification. This decreased proteolytic fragmentation to a minimum and lead to a single species sample that was subsequently used for crystallization (Fig. 5.3).

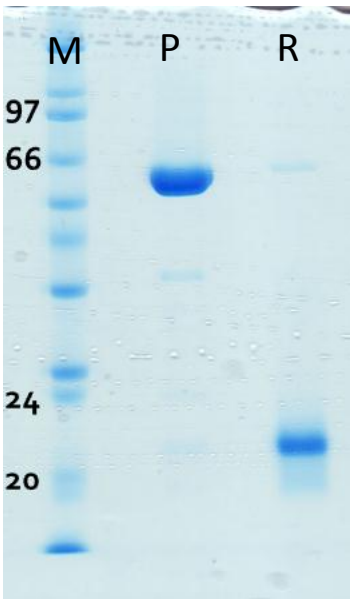


Figure 5.3: SDS-PAGE of samples used for crystallization. M) Marker, molecular weight shown in kDa, P) Plexin-$B1_{cyto}$, R) Rac1 Q61L

Previous studies have shown that Plexin-B1 interacts with the small GTPases Rnd1, Rac1 and RhoD in their active, GTP-bound state [194]. Rac1 was rendered constitutively active by introducing a Gln61Leu mutation [202] in addition to loading with the non-hydrolyzable GTP analogue GppNHp. This Rac1 mutant, expressed in *E. coli*, was used in all subsequent experiments and is named Rac1* hereafter.

The dissociation constant for all three GTPases and the Plexin-RBD has been reported to be in the low micromolar range [194]. However, the complex between Plexin-$B1_{cyto}$ and Rac1* could not be purified in a quantitative manner on a gelfiltration column (Fig. 5.2).

5.4 Crystallization and structure determination of the Plexin-B1-Rac1* complex

All three plexin constructs Plexin-$B1_{cyto}$, Plexin-$B1_{\Delta1}$ and Plexin-$B1_{\Delta2}$ were expressed in insect cells (see Section 2.2.2) mixed with Rac1* in a molar ratio of 1 : 1.1 and crystallization trials were set-up at a protein concentration of 10mg/ml in 10 mM Hepes, pH 7.5, 150 mM NaCl, 2 mM TCEP. The Plexin-$B1_{cyto}$-Rac1* complex crystallized initially in 1 M Li_2SO_4, 0.5 M $(NH_4)_2SO_4$ and 100mM Citrate, pH 5.6, at 20°C. Several rounds of optimization, in which the precipitant concentration, the pH, the protein to reservoir drop ratios and a variety of different additives were explored, finally lead to crystals resembling thin squares and a complete dataset was collected at beamline I03 at Diamond Light Source, UK. The Plexin-$B1_{\Delta1}$-Rac1* complex crystallized in 20% PEG 3350, 0.2M KSCN, 0.1M Bis-Tris Propane, pH 6.5. After a similar optimization strategy as applied for the former complex Plexin-$B1_{\Delta1}$-Rac1* crystallized as thin needles and data were collected at the microfocus beamline ID23-2 at the European Synchrotron Radiation Facility, France, following a helical data collection strategy. In this data collection strategy the crystal is translated after the collection of a number of frames, in this case 30. This ensures that a complete dataset can be collected from the same crystal while at the same time optimizing and maximizing the high-resolution limits by longer beam exposure. Both structures were solved by molecular replacement using PHASER [122] with the structure of human Plexin-B1 (PDB ID: 3HM6 [44]) and active Rac1 (PDB ID: 1MH1 [203]) as search model. In both cases the molecular replacement solution showed high Z-scores above 7, positive log-likelihood gains and a clearly discernible electron density. In particular, the GTP binding site of the RhoGTPase showed positive density for the nucleotide that was not included in the molecular replacement search model. The solution

was manually adjusted using COOT [124] and refined using autoBUSTER [125]. Refinement statistics are given in Table 3; all data within the indicated resolution range were included. The 4.2 Å structure was refined using three-fold NCS as implemented in autoBUSTER [125] and tight geometric restraints to minimize the introduction of any model bias. Data collection and refinement statistics can be found in Table 3.

Stereochemical properties were assessed by MOLPROBITY [127]. Ramachandran statistics are as follows (favored/disallowed (%)):Plexin-B1$_{cyto}$-Rac1* 91.7/0.2, Plexin-B1$_{\Delta 1}$-Rac1* 95.5/0.2 (pre-proline residue Leu1981 is in a disallowed region in both structures). Superpositions were calculated using SHP[204]. Buried surface areas of protein-protein interactions were calculated using the PISA webserver (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html).

Table 3. Data collection and refinement statistics

	Plexin-B1$_{\Delta 1}$ – Rac1* complex	PlexinB1$_{cyto}$ – Rac1* complex
Data collection		
Resolution (Å)	3.10 (3.21-3.10)[a]	4.20 (4.35-4.20)
Space group	C2	$I2_12_12_1$
Cell dimension		
a, b, c (Å)	184.1, 63.9, 84.6	142.1, 224.1, 258.3
α, β, γ (°)	90, 107.5, 90	90, 90, 90
Redundancy	5.3 (5.3)	5.1 (5.1)
Completeness (%)	100 (100)	96.0 (97.8)
R_{merge} (%)	16.0 (91.4)	9.5 (83.9)
Avg I/σ	10.4 (1.7)	13.7 (1.3)
Refinement		
Resolution (Å)	46.0 – 3.1	42.0 – 4.2
No. Reflections	17,340 (2,670)	29,381 (2,832)
R_{work}/R_{free}[b] (%)	21.4 (26.6)/ 24.7 (35.2)	24.6 (27.5)/ 26.9 (29.2)
Complexes in the asymmetric unit	1	3
No. Atoms		
Protein	5,762	16,493
B-factors		
Protein (Å^2)	67.9	123.3
r.m.s. Deviations		
Bond lengths (Å)	0.008	0.009
Bond angles (°)	0.94	1.09

[a] Numbers in parenthesis are for the highest resolution shell

[b] R_{free} equals the R_{work} against 5% of the data removed prior to refinement.

5.5 Structure of the intracellular domain of Plexin-B1 in complex with constitutively active Rac1

The crystal structure of Plexin-B1$_{\Delta 1}$ in complex with Rac1* was determined to a resolution of 3.1 Å and refined to a crystallographic R-factor of 21.4% (R_{free} = 24.7%, Fig. 5.4, Table 3, Fig. 5.5).

The overall structures of Plexin-B1$_{\Delta 1}$ and Rac1* in the complex are very similar to their apo-structures [44,203], with rmsd values of 1.5Å and 0.6Å, respectively. However, there is some flexibility between the Plexin-B1 GAP and the RBD with the RBD being rotated by ~6° compared to the apo-structure (Fig. 5.6). Rac1* binds exclusively to the RBD and does not form any contacts with the GAP domain. The interface between Rac1* and the Plexin-B1 RBD covers a buried surface area of 707Å^2 and is dominated by hydrophobic interactions. Plexin-B1 residues Trp1807^{Plex}, Leu1815^{Plex}, Thr1823^{Plex} and Tyr1839^{Plex} form a continuous hydrophobic patch that is complemented by Rac1 residues Phe37^{Rac}, Val36^{Rac}, Leu67^{Rac} and Leu70^{Rac} (Fig. 5.4). All of these residues are almost completely buried within the interface (at least 80% of the solvent accessible surface area) with the exception of Val36^{Rac} (38%). Thr1823^{Plex} and Tyr1839^{Plex} are part of a potential hydrogen bonding network involving Asp1821^{Plex}, Ser1824^{Plex}, Asn1834^{Plex} and His1838^{Plex} that is likely to be crucial for the structural integrity of the domain. The hydrophobic interaction between Plexin-B1$_{\Delta 1}$ and Rac1* is extended by two potential hydrogen bonds formed between the sidechain of Asp38^{Rac} and the backbone amides of Val1811^{Plex} and Ala1812^{Plex}. Remarkably, all of the Plexin-B1 residues described above are conserved across A- and B-class plexins (Fig 5.7) therefore most likely preserving this mode of recognition.

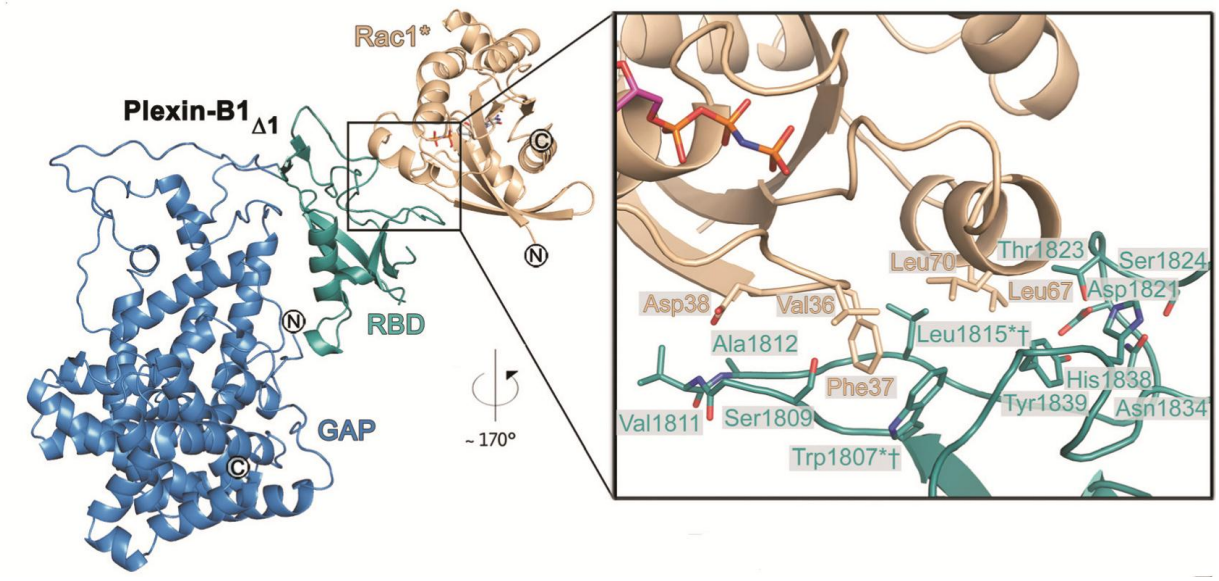


Figure 5.4: Structure of the monomeric Plexin-B1$_{\Delta 1}$-Rac1* complex. Ribbon diagram of the complex between Plexin-B1$_{\Delta 1}$ and Rac1* with colour coding for Plexin-B1 as in Fig. 5.1 and Rac1* in wheat. GppNHp and residues involved in the interaction are shown in stick representation. Residues mutated in SPR binding experiments are marked with an asterisk, mutations analysed in the cell collapse assay with a blade.

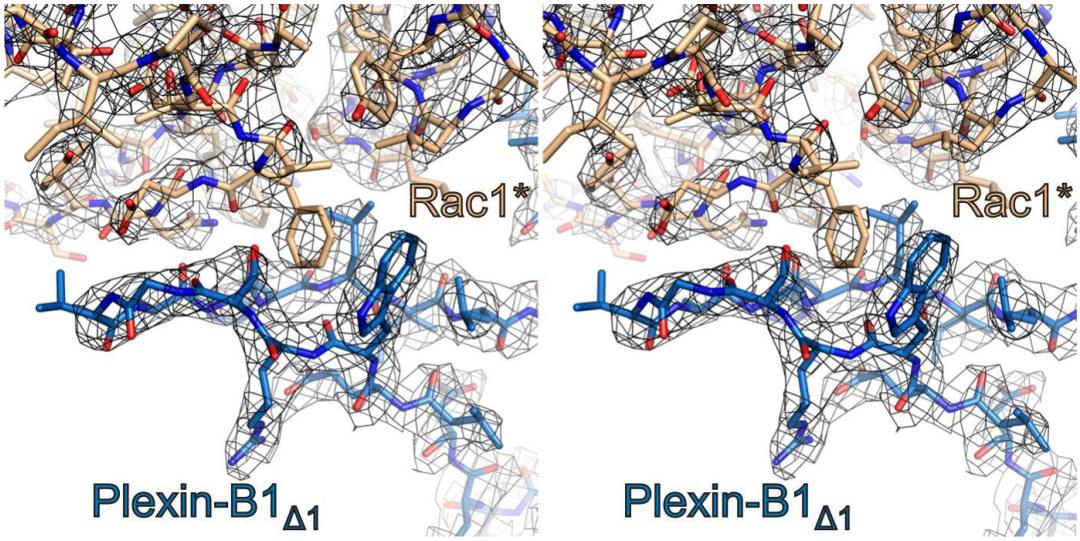


Figure 5.5: Stereoview of the electron density of the Plexin-B1Δ1-Rac1* interface. The orientation is similar to Fig.1b, right panel. The density represents a 3.2 Å SigmaA-weighted 2*Fobs-Fcalc* map contoured at 1.0σ.

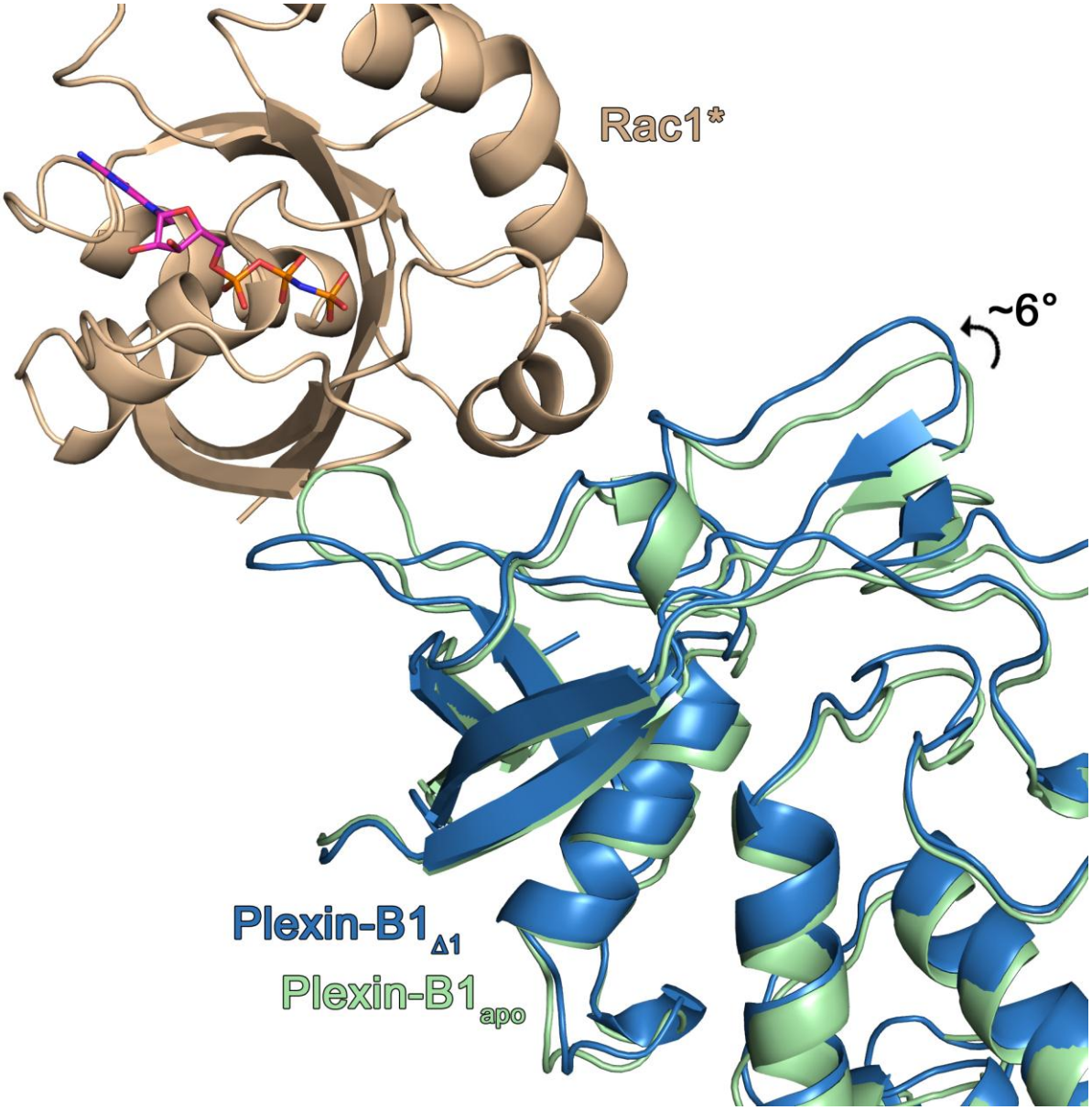


Figure 5.6: Superposition of the Plexin-B1 apo structure onto the Plexin-B1$_{\Delta 1}$-Rac1* complex. Colour coding is as in Fig. 1b with Plexin-B1apo in palegreen. Coordinates for the Plexin-B1apo structure can be found under PDB ID: 3HM6. The complexes were aligned onto the plexin GAP domains using SHP. The orientation is similar to Fig.1b, right panel. The slight rotation of the Plexin-B1Δ1 RBD in comparison to the Plexin-B1apo structure is indicated by an arrow.

Plexin_B1_human
Plexin_B1_human
Plexin_A3_mouse
Plexin_A_fly
Plexin_B_fly
Plexin_D1_human
Plexin_C1_human

Figure 5.7: Sequence alignment of the intracellular region of plexins from different classes and organisms. The plexin 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 Plexin-B1. Secondary structure elements are shown for human Plexin-B1. Residues that were mutated and studied in SPR as well as in the cellular collapse assays are marked with a dot, those that were only studied in the cellular assay are marked with a triangle.

On Rac1*, all residues mentioned above map onto the switch I or switch II region [58] (Figure 5.8) whose conformation resembles that of active Rac1 in other Rac1-effector complexes [205]. Since these regions undergo large conformational changes upon GTP binding, this explains why Plexin-B1 is highly specific for active, GTP-bound Rac1 [194]. Recently, the structure of the RBD fragment of Plexin-B1 in complex with the constitutively active RhoGTPase, Rnd1, has been reported [44]. Structural superposition of the RBD-RhoGTPase complexes gives an rmsd of 0.96Å (Figure 5.9). Despite a sequence identity of only 32% between Rac1 and Rnd1, the Plexin-B1 RBD-Rnd1 complex interface is very similar to the one described here. All hydrophobic interactions as well as the two potential hydrogen bonds are conserved in both structures.

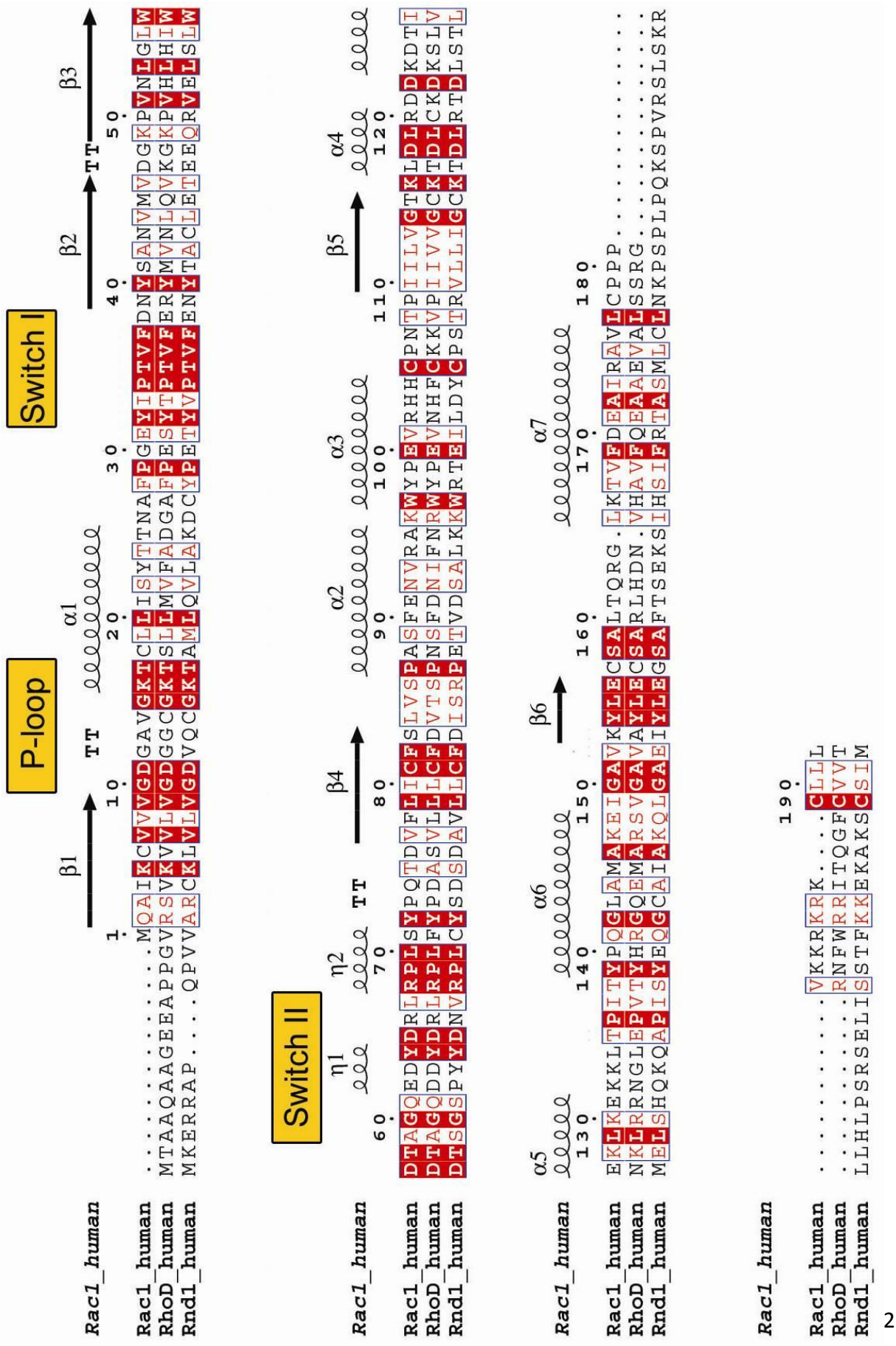

P-loop
Switch I
Switch II
Rac1_human
Rac1_human
RhoD_human
Rndl_human

Figure 5.8: Sequence alignment of the human RhoGTPases Rac1, Rnd1 and RhoD. The alignment is prepared as described in Supplementary Figure 2. Numbering corresponds to human Rac1. Secondary structure elements are shown for human Rac1. The three regions characteristic for small GTPases and their activation state, the P-loop, switch I and switch II, are marked by yellow boxes.

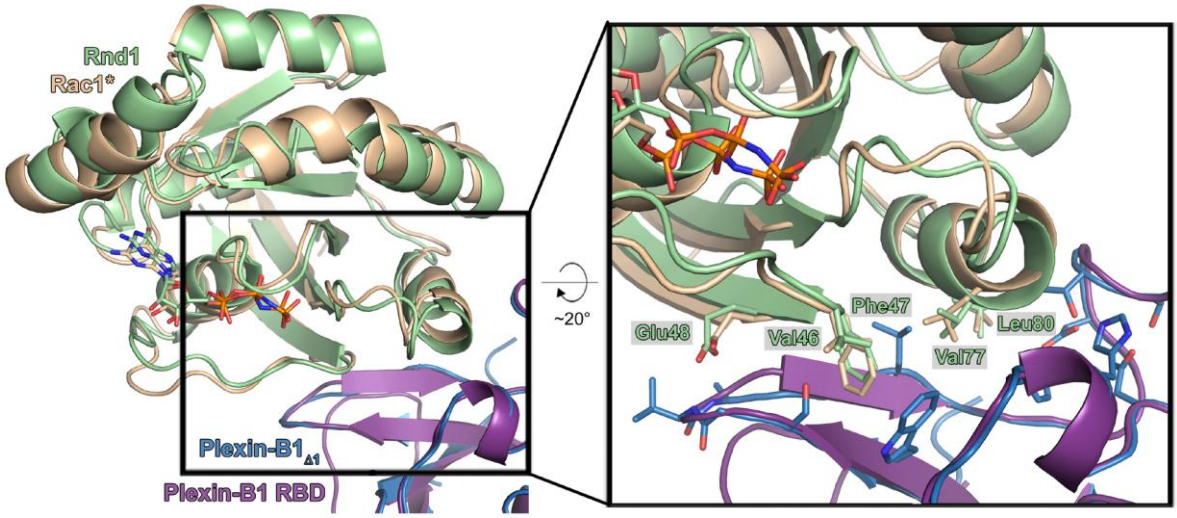


Figure 5.9: Superposition of the Plexin-B1 RBD-Rnd1 complex onto the Plexin-B1$_{\Delta 1}$-Rac1* structure. Colour coding is as in Fig. 5.3b. Rnd1 is in palegreen and the Plexin-B1 RBD of the Plexin-B1 RBD-Rnd1 complex in purple. Coordinates for the Plexin-B1 RBD-Rnd1 complex can be found under PDB ID: 2REX. The complexes were aligned onto the plexin molecules using SHP. The orientation is similar to Fig. 5.3b, right panel, with Rnd1 residues labelled in palegreen.

To corroborate our structural findings the affinity between Plexin-B1$_{cyto}$ and Rac1*, as well as Rnd1, was studied using SPR. Rnd1 is constitutively active due to its lack of GTPase activity [206]. Plexin-B1$_{cyto}$ binds to Rac1 and Rnd1 with an affinity of 18.9µM and 22.9µM, respectively (Fig. 5.10, Table 4), which is in agreement with recently published affinities determined by isothermal titration calorimetry [44]. A series of Plexin-B1 mutations in the hydrophobic interface, Trp1807Glu^{Plex}, Leu1815Pro^{Plex} (previously linked to prostate cancer [188]) and Leu1815Glu^{Plex}, completely abolished its interactions with Rac1* and Rnd1 (Fig. 5.10).

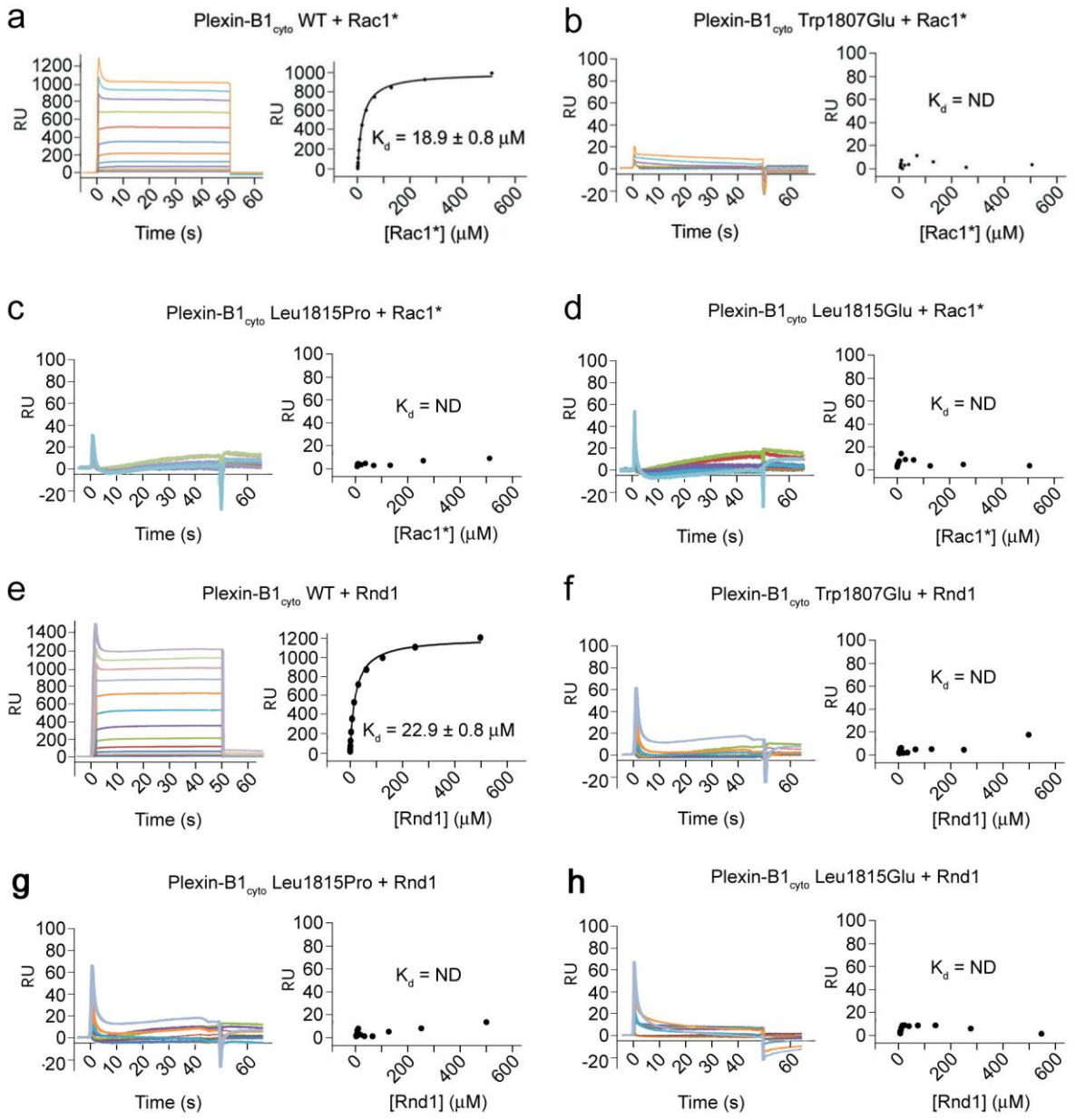


Figure 5.10: Binding of Rac1* and Rnd1 to site A mutants of Plexin-B1. Left, representative sets of experimental sensorgrams from typical equilibrium-based binding experiments, with reference subtraction. Different concentrations of the respective RhoGTPase were injected over surfaces coupled with the plexin constructs. For all injections, the experimental traces reached equilibrium and returned to baseline after the injection. Right, plot of the equilibrium binding response (response units (RU)) against RhoGTPase concentration ranging from 120 nM to 500 µM. Within one experiment each concentration was measured twice. All experiments were performed in duplicate. Best-fit binding curves corresponding with a 1:1 binding model are shown as lines. Binding constants (K_d) are given as mean with the error representing the standard error of the mean. WT, wild-type; ND: not determinable. (a) Plexin-$B1_{cyto}$ Leu1815Pro + Rac1*, (b) Plexin-$B1_{cyto}$ Leu1815Glu + Rac1*, (c) Plexin-$B1_{cyto}$ WT + Rnd1, (d) Plexin-$B1_{cyto}$ Trp1815Glu + Rnd1, (e) Plexin-$B1_{cyto}$ Leu1815Pro + Rnd1, (f) Plexin-$B1_{cyto}$ Leu1815Glu + Rnd1.

K_d in µM	WT	Trp1807Glu	Leu1815Pro	Leu1815Glu
Rac1*	18.9 ± 0.8	-	-	-
Rnd1	22.9 ± 0.8	-	-	-

Table 4: Binding constants (K_d) measured by SPR between different Plexin-B1$_{cyto}$ mutants and Rac1* and Rnd1. Data are expressed as mean ± standard error and no detectable binding is represented by a dash.

To validate these effects on binding in a functional context, COS cell-based collapse assays were performed with the full-length transmembrane receptor, testing for Plexin-B1 activity *in vivo* [129]. Surprisingly, none of the mutants shown to abolish Rac1* or Rnd1 binding had an effect on the collapse response of the cells (Fig. 5.11). Since the necessity of RhoGTPase binding for plexin function is well established [49,197], it was unclear how to correlate the biophysical and cellular results.

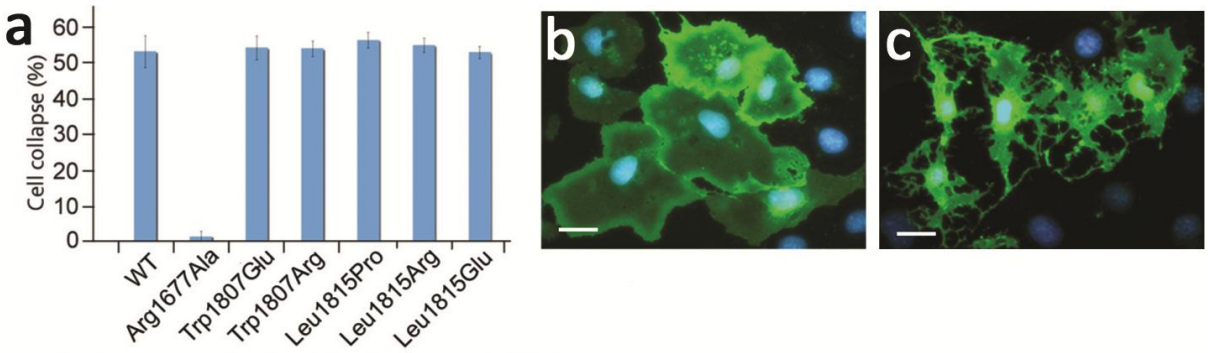


Figure 5.11: COS-7 cell collapse assay. (a) Histogram showing the effect of different mutations on semaphorin induced collapse. Each experiment was repeated twice and 2 × 200 cells were counted each time. Results represent the mean with error bars showing standard error. (b-c) Representative images of non-collapsed cells (b) and SEMA4D-induced collapsed cells (c). Scale bar, 40 µm. Arg1667Ala is a constitutively inactive Plexin-B1 mutant and served as negative control.

5.6 Three-fold arrangement of the Plexin-B1$_{cyto}$-Rac1 unit

The relative position of Rac1* in regard to the putative Ras binding site revealed no mechanism for the direct regulation of the catalytic activity of Plexin-B1 by the small RhoGTPase. To address whether the N-terminal residues missing in the Plexin-B1$_{\Delta 1}$ construct might harbour an important site for RhoGTPase mediated plexin activity, the crystal structure of the entire cytoplasmic domain of Plexin-B1 (Plexin-B1$_{cyto}$) in complex with Rac1* (Fig. 5.13) was solved. The 4.2 Å-model is of high-quality for this resolution range, reflected by the crystallographic R-factor of 24.6% (R_{free} = 26.9%, Table 3, Fig. 5.14). The asymmetric unit contains a trimeric arrangement comprising three copies of the Plexin-B1$_{cyto}$-Rac1* unit, with each Rac1* molecule contacting two Plexin-B1$_{cyto}$ molecules (Fig. 5.13a-b). This arrangement is not the result of crystallographic symmetry but does show near perfect three-fold geometry (120° between pairs of Rac1* molecules and 117°, 119° and 124° between the copies of Plexin-B1$_{cyto}$). Moreover, the interfaces between Plexin-B1$_{cyto}$ and Rac1* are essentially identical across the three copies in the asymmetric unit and are not found in any of the crystallographic symmetry generated interfaces. These observations strongly suggest that this three-fold complex is not purely a product of crystal lattice formation. It was not possible to show a three-fold complex with the soluble constructs in analytical ultracentrifugation at a plexin concentration of 250 µM (Fig. 5.15). This suggests that high local concentrations in the crystal or indeed at the plasma membrane are necessary for this arrangement and that the three-fold interaction might be too weak to be detected in solution [207].

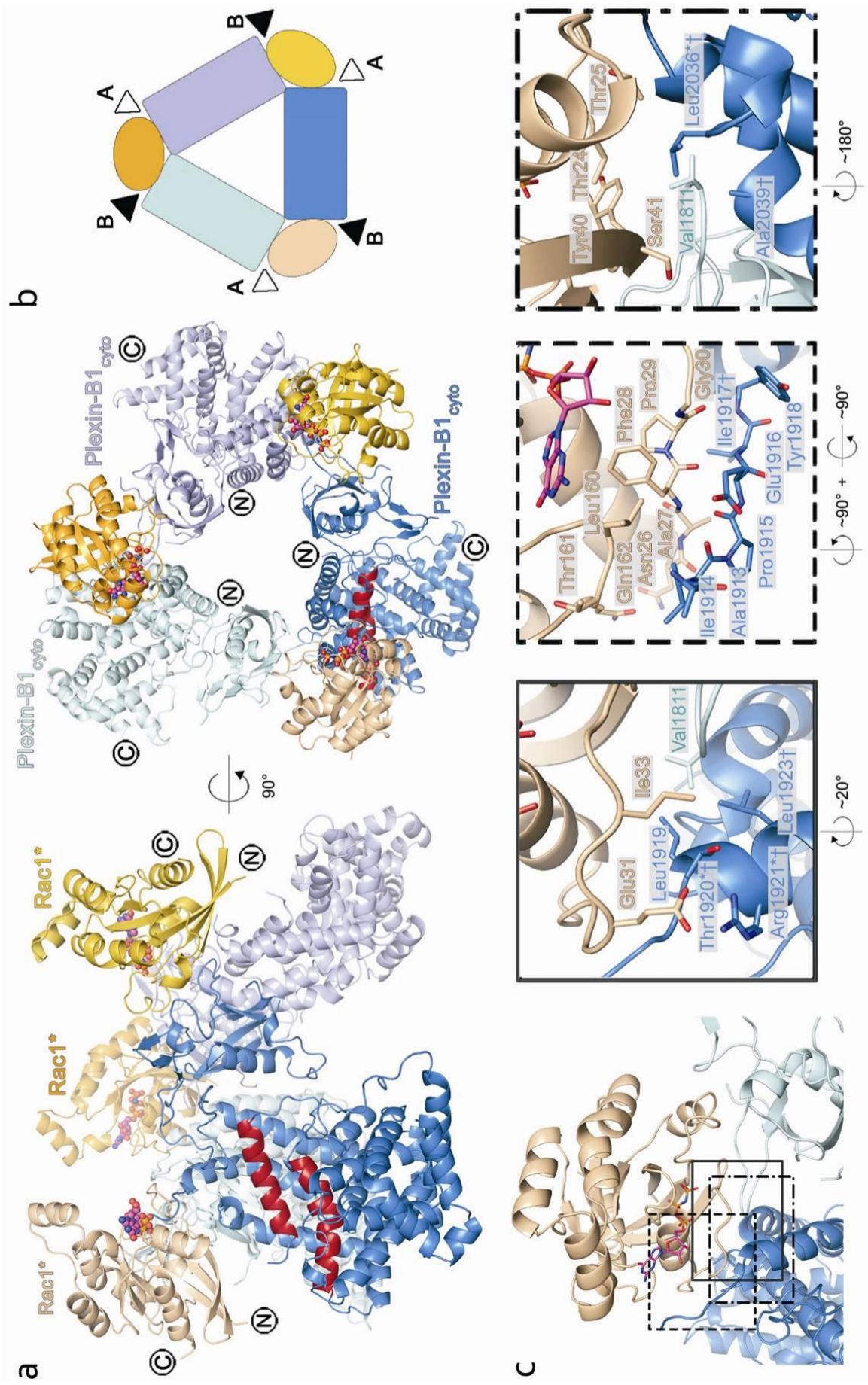

a
Rac1*
Rac1*
Rac1*
C
N
90°
Plexin-B1cyto
Plexin-B1cyto
Plexin-B1cyto
b
A
B
A
B
A
B
C
Glu31
Ile33
Val1811
Leu1919
Thr1920*†
Arg1921*†
Leu1923†
~20°
Thr161
Leu160
Gln162
Asn26
Ala27
Phe28
Pro29
Gly30
Ile1914
Ala1913
Pro1915
Ile1917†
Glu1916
Tyr1918
~90° + ~90°
Tyr40
Thr24
Thr25
Ser41
Val1811
Leu2036*†
Ala2039†
~180°

Figure 5.13: Structure of the three-fold Plexin-B1$_{cyto}$-Rac1 complex. (a) Overview of the three-fold arrangement of the three Plexin-B1$_{cyto}$-Rac1* units. Rac1* is coloured in wheat, orange and yellow and Plexin-B1$_{cyto}$ is coloured in skyblue, paleblue and palecyan, respectively. GppNHp is shown as spheres. Termini for Rac1* are labelled in the left panel, and termini for Plexin-B1$_{cyto}$ in the right panel. α-helices 13, 14 and 15 harbouring the putative Ras-binding site are highlighted in red. (b) Schematic overview of the complex. The orientation and colour coding is as in (a), right panel. The two Rac1 binding sites, site A and site B, are marked by arrowheads. (c) Detailed view of the Rac1 binding site B. The orientation of the left panel is similar to Fig. 5.3b, right panel. The three boxed panels correspond to the close-up views depicted in the left panel. Colour coding is as in (a).

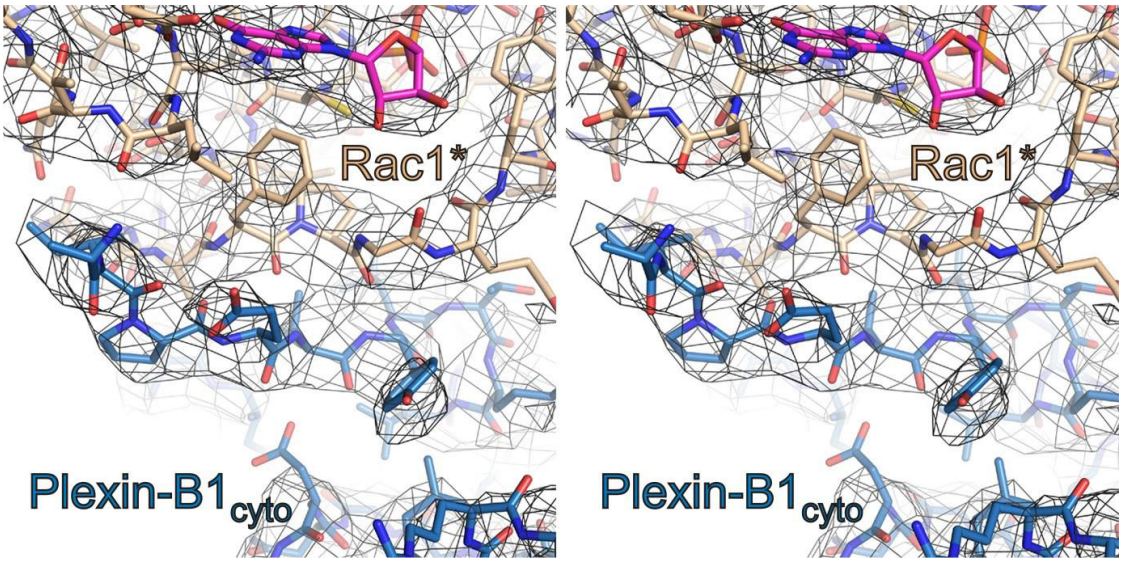


Figure 5.14: Stereoview of the electron density of the Plexin-B1$_{cyto}$-Rac1* site B interface. The orientation is similar to Fig.2b, third panel. The density represents a 4.2 Å SigmaA-weighted $2F_{obs}$-F_{calc} map contoured at 1.0σ.

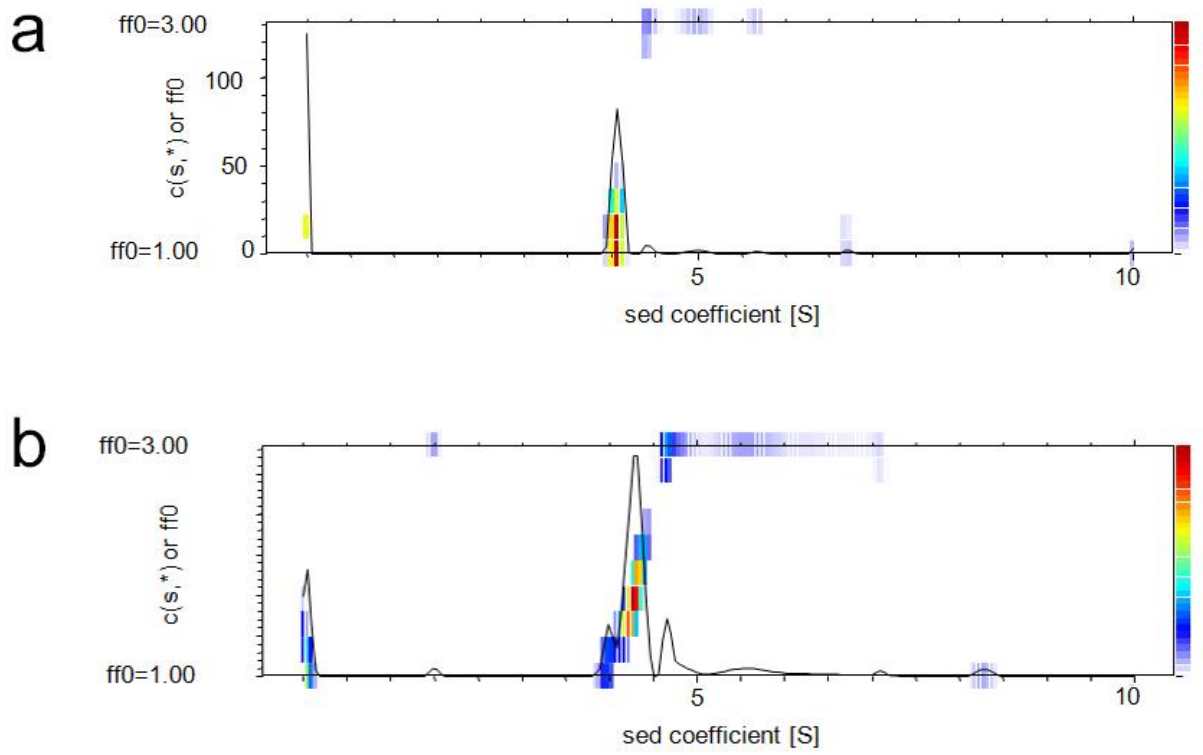


Figure 5.15: Analysis of sedimentation velocity experiments with Plexin-B1 alone (a) and in complex with Rac1 (b). Both experiments were carried out at a protein concentration of 25 mg/ml. The calculated sedimentation coefficient for the trimer is 10.2.

There are no major conformational changes within the three Plexin-$B1_{cyto}$-Rac1* units when compared to the one-to-one complex. This three-fold arrangement is mediated by a previously unidentified second binding site for the Rho-GTPase on Plexin-$B1_{cyto}$ (site B, the previously observed Plexin-B1 RBD-RhoGTPase interface is hereafter called site A). Site B involves the N-terminal region of α-helix 11 and the loop preceding it (residues 1913^{Plex}-1923^{Plex}) plus α-helix 16 (residues 2036^{Plex}-2039^{Plex}) and is in close proximity to the putative Ras binding site (Fig. 5.7, Fig. 5.13). It covers a total buried surface area of ~570$Å^2$, therefore significantly extending the interface for Rac1 binding, and is dominated by hydrophobic interactions. On Plexin-$B1_{cyto}$ the majority of contacts (60% of buried surface area) are made

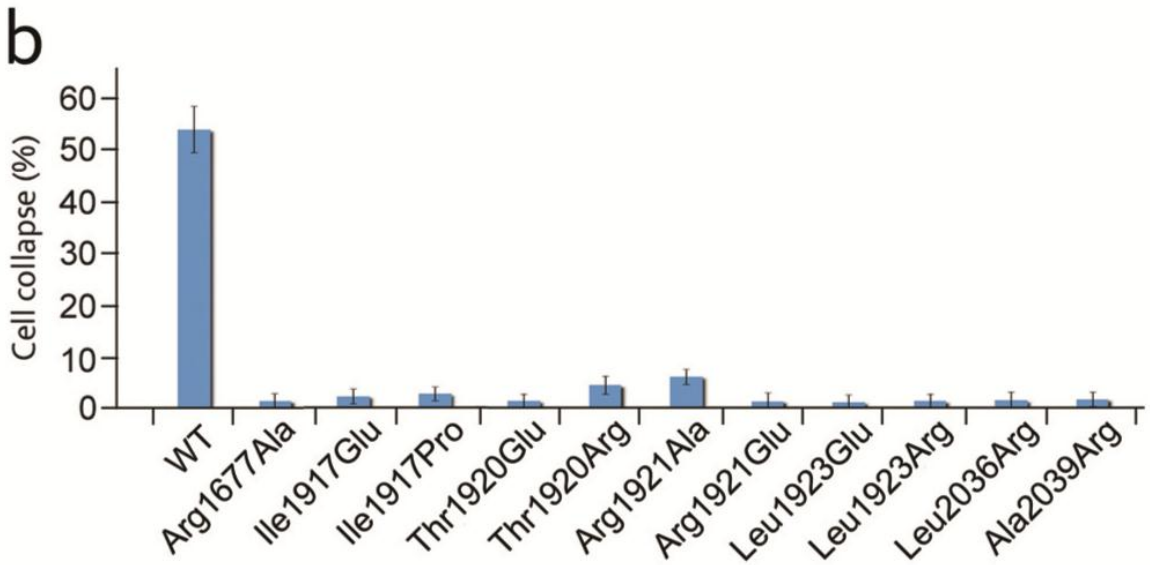
by the loop residues 1913^{Plex}-1918^{Plex} (Fig. 5.13c). Interestingly, these residues are disordered in the apo-structures of Plexin-B1 and Plexin-A3. The site B interface on Rac1* is predominantly formed by residues that precede the switch I region (residues 24-33, Fig. 5.13c). The conformation of these residues is known not to depend on the activation state of the RhoGTPase [58], thus the specificity of Plexin-B1 for active RhoGTPases appears to result exclusively from interactions with site A. The three-fold complex is further stabilized by contacts between two adjacent Plexin-$B1_{cyto}$ molecules, on the one side mainly involving a loop comprising residues 1808^{Plex}-1813^{Plex}, and on the other side a surface directly adjacent to site B (residues 1919^{Plex}-1938^{Plex} and residues 2036^{Plex}-2044^{Plex}, Fig. 5.13c). However, this plexin-plexin interaction is unlikely to be stable without the addition of the bridging Rac1* since it only contributes a total buried surface area of ~310 $Å^2$.

In order to assess the potential functional significance of site B, three Plexin-B1 mutants, $Thr1920Glu^{Plex}$, $Arg1921Ala^{Plex}$ and $Leu2036Arg^{Plex}$ were designed. Firstly, the binding affinity of these mutants to Rac1* and Rnd1 was studied using SPR. None of the site B mutations had a significant effect on the affinity towards either of the RhoGTPases, suggesting that site A alone is sufficient for Rac1* and Rnd1 binding (Fig. 5.16, Fig. 5.17). However, these mutations as well as additional ones at these and other site B residues ($Ile1917^{Plex}$, $Leu1923^{Plex}$ and $Ala2039^{Plex}$) completely abolished the typical collapse response in the COS-cell assay (Fig. 5.16b). Every site B mutation tested was detrimental to Plexin-B1 activity and led to a complete loss of function. All mutant proteins showed a similar expression level, as judged by immunofluorescence, and were present in the plasma membrane, indicating their structural integrity. This finding was explored further in an independent experimental assay

to monitor directly Ras GTPase activity in vivo. In agreement with my results from the collapse assay, the Plexin-B1 site B mutants also did not show any GAP activity towards R-Ras (Fig. 5.18). These results indicate that although site B is not essential for binding of the RhoGTPase it is crucial for Plexin-B1 activity suggesting that the three-fold complex seen in the crystal has functional significance. In accordance with this putative functional role, all of the residues in site B are conserved across all species and classes of plexins with the exception of Ala1913^{Plex} and Pro1915^{Plex} whose sidechains do not participate in the Rac1-site B interaction (Fig. 5.13c).

a

K_d in µM	Thr1920Glu	Leu2036Arg	Arg1921Ala
Rac1*	21.0 ± 0.4	11.9 ± 1.5	12.0 ± 0.5
Rnd1	14.2 ± 0.4	16.5 ± 2.3	15.2 ± 3.7



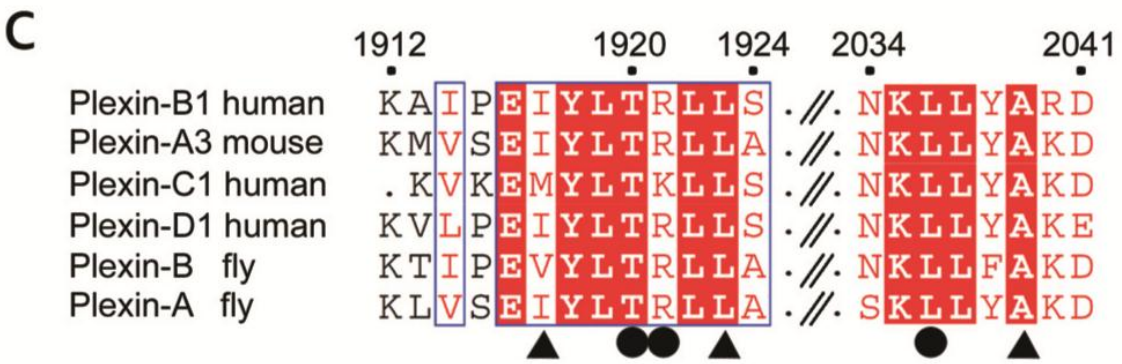


Figure 5.16: Characterization of the second binding site. (a) Binding constants (K_d) measured by SPR between different Plexin-B1 constructs carrying mutations in binding site B, and Rac1* and Rnd1, respectively. Data are expressed as mean ± standard error. (b) Histogram showing the effect of different site B mutations on semaphorin-induced collapse. Results are shown as mean with error bars representing standard error of the mean. (c) Sequence alignment of residues involved in site B interaction. Residues that were mutated in the cell collapse assay are marked with a triangle and those that were analysed in both SPR and cellular assays are marked with a dot.

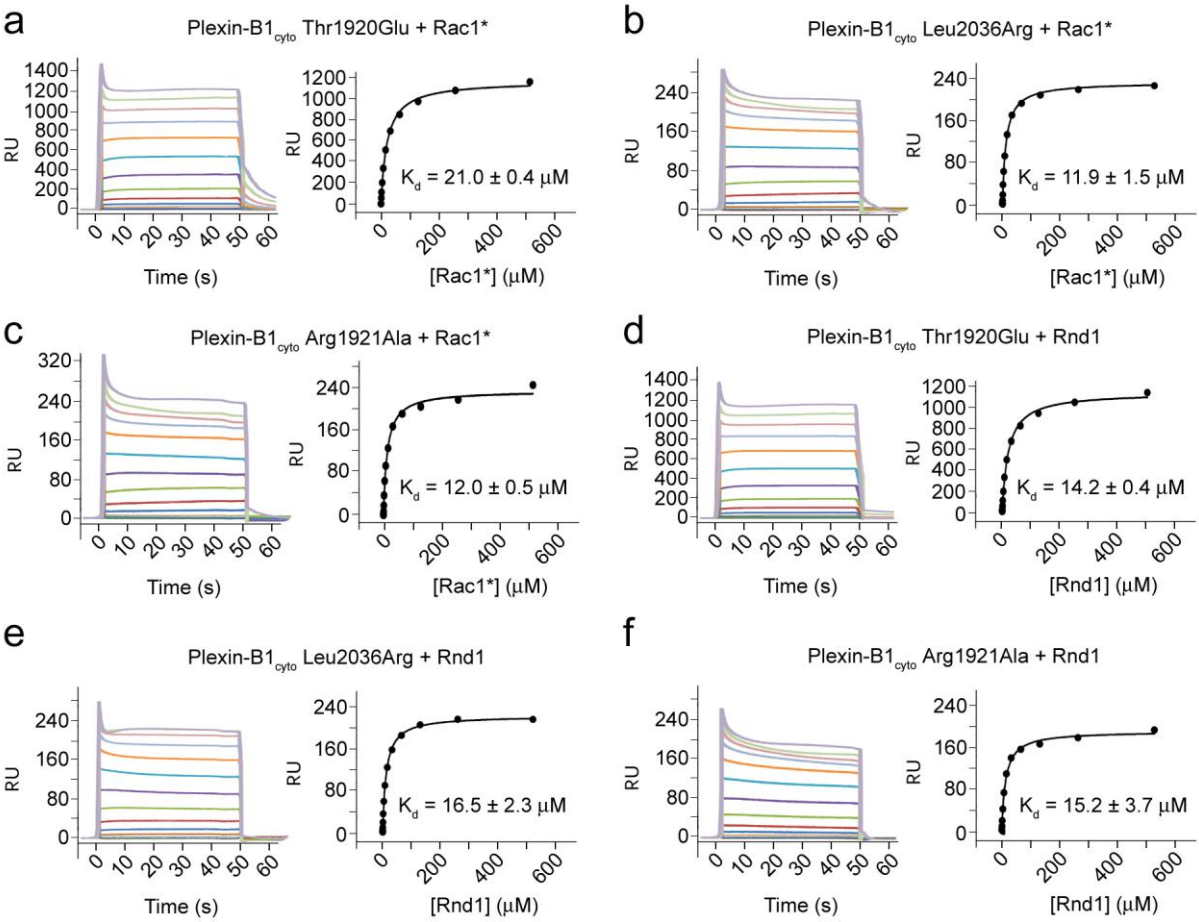


Figure 5.17: Binding of Rac1* and Rnd1 to site B mutants of Plexin-B1. Data are presented as in Figure 5.9 (a) Plexin-B1$_{cyto}$ Thr1920Glu + Rac1*, (b) Plexin-B1$_{cyto}$ Leu2036Arg + Rac1*, (c) Plexin-B1$_{cyto}$ Arg1921Ala + Rac1*, (d) Plexin-B1$_{cyto}$ Thr1920Glu + Rnd1, (e) Plexin-B1$_{cyto}$ Leu2036Arg + Rnd1, (f) Plexin-B1$_{cyto}$ Arg1921Ala + Rnd1.

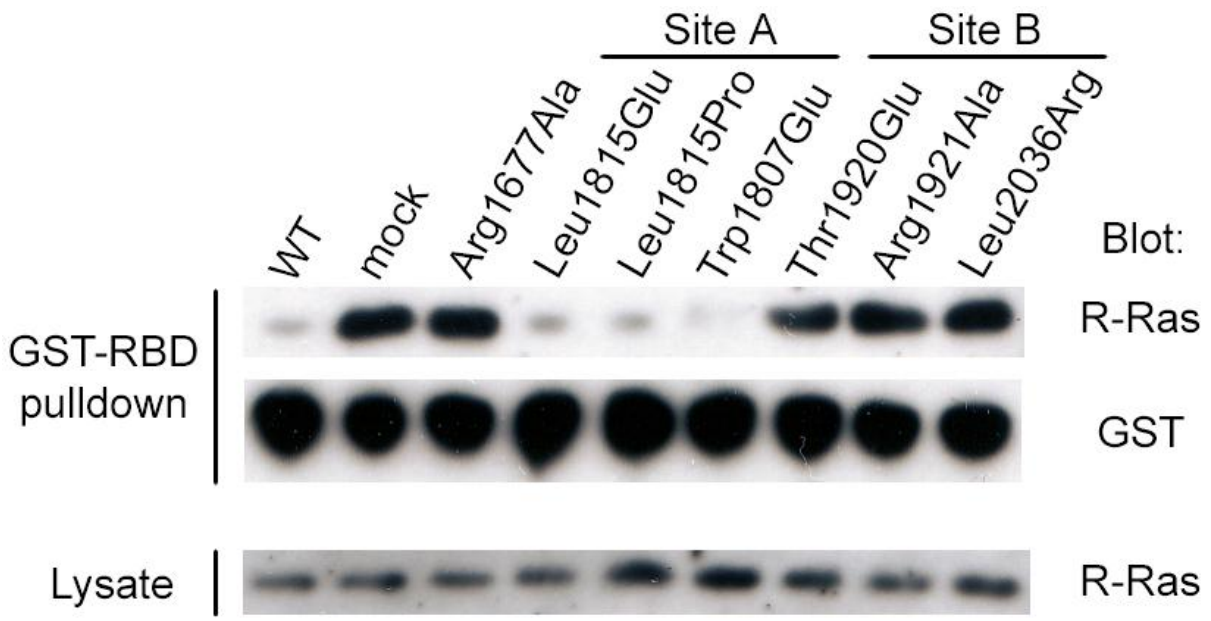


Figure 5.18: Mutations in site B but not site A abolish Plexin-B1 RasGAP activity. COS-7 cells transfected with full-length Plexin-B1 and its mutants were stimulated with $SEMA4D_{ecto}$ for 10min. The cell lysates were incubated with GST-fused Ras-binding domain of Raf-1 and bound R-Ras and total cell lysates were detected by immunoblotting. The results shown are representative of two independent experiments that yielded similar results. WT, wild-type; mock, chicken receptor protein tyrosine phosphatase σ Ig1-2.

5.7 The importance of the N-terminal segment of the Plexin-B1 cytoplasmic region

The three-fold complex revealed by the crystal structure and the cell collapse assays suggest that GTPase binding at site B contributes to plexin function. However, the SPR experiments reveal no direct evidence for GTPase binding at this site. Therefore an explanation for this lack of binding was sought. In both the one-to-one and three-fold complex structures it was not possible to trace the N-terminal helix (residues 1511^{Plex}-1562^{Plex}, Fig. 5.7) due to a lack of well-ordered electron density. In the apo-structure of mouse Plexin-A3 [43] this region was well-defined. Superposition of the Plexin-A3 structure with the three-fold complex reveals that this helix would block site B, therefore preventing its interaction with Rac1*

(Fig. 5.19). This steric hindrance model predicts that shortening of the N-terminal helix will remove this block and allow Rac1* and Rnd1 to bind to site B.

To test this model mutant constructs lacking the N-terminal helix (Plexin-B1$_{\Delta 2}$) were generated and assayed for RhoGTPase binding in SPR. Indeed, RhoGTPase binding to the site A Plexin-B1$_{cyto}$ mutant Leu1815Glu^{Plex} was now observed, suggesting that truncation of the N-terminal helix has exposed site B (Fig. 5.19b, Fig. 5.20). However, binding can only be observed with high coupling densities of the Plexin-B1$_{\Delta 2}$ on the SPR chip (Fig. 5.19b). This is consistent with a bivalency effect in which two adjacent plexin molecules bind the same RhoGTPase molecule implying that the mutated site A is still competent to contribute to an avidity effect [38]. At low coupling densities the Plexin-B1$_{\Delta 2}$ molecules are too far apart from each other to allow a bivalent interaction with Rac1* or Rnd1. There was no increase in affinity for wild-type N-terminal truncated Plexin-B1$_{\Delta 2}$ compared to the full-length Plexin-B1$_{cyto}$ even at high coupling densities (data not shown). Binding studies on the isolated Plexin-B1 RBD show similar affinities [194] to those that were determined for the full-length cytoplasmic region. Thus these observations suggest that, if intact, site A dominates binding to the RhoGTPase.

Interestingly, sequence analysis of the N-terminal cytoplasmic segment of Plexin-B1 (residues 1511-1539) predicts a trimeric coiled-coil (Fig. 5.21) and similar regions in other plexins from all classes are also predicted to adopt a coiled-coil conformation. In accordance with this, Ile1563^{Plex} in Plexin-B1$_{cyto}$, the first residue visible in the electron density map, points towards the inside of the three-fold complex locating the three N-terminal segments in close proximity to each other (Fig. 5.13a, right panel). This proximity suggests an

explanation for the observation of the higher oligomeric state three-fold complex in the crystal. Although the N-terminal helix is not well-ordered in the structure, it could form a trimeric coiled-coil albeit containing significant flexibility.

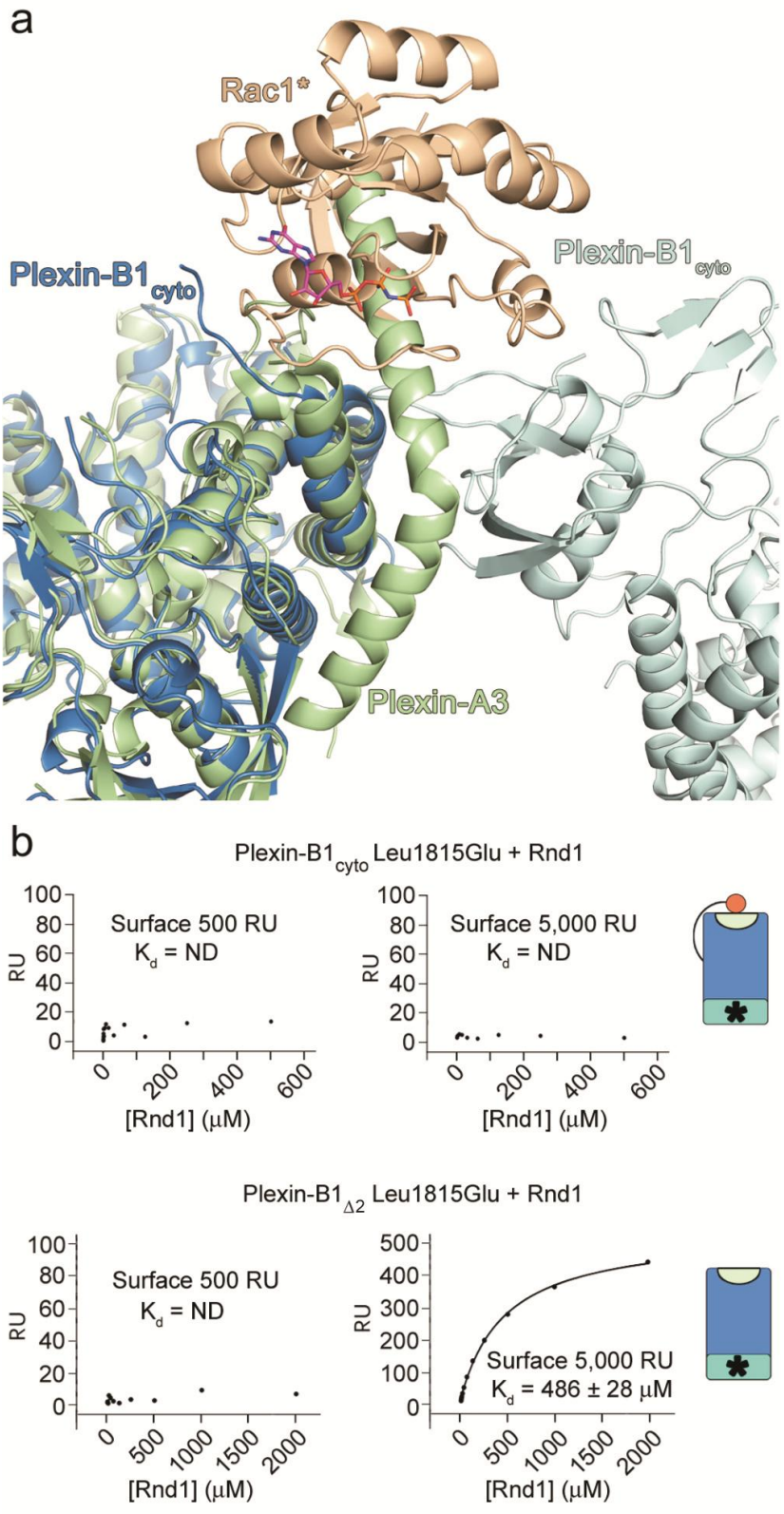


Fig. 5.19: Analysis of the Plexin-B1$_{cyto}$ N-terminal juxtamembrane helix. (a) Superposition of mouse Plexin-A3 (PDB ID: 3IG3) onto the three-fold Plexin-B1$_{cyto}$-Rac1* complex. Colour coding is as in Fig. 5.13a. Plexin-A3 is shown in palegreen. The juxtamembrane helix ordered in the Plexin-A3 structure would block Rac1* binding to site B. (b) Representative plots of the equilibrium binding response against Rho GTPase concentration ranging from 120 nM to 2,000 µM. ND: not determinable. The plexin constructs used in each experiment are schematically presented on the right with colour coding as in Fig. 5.22.

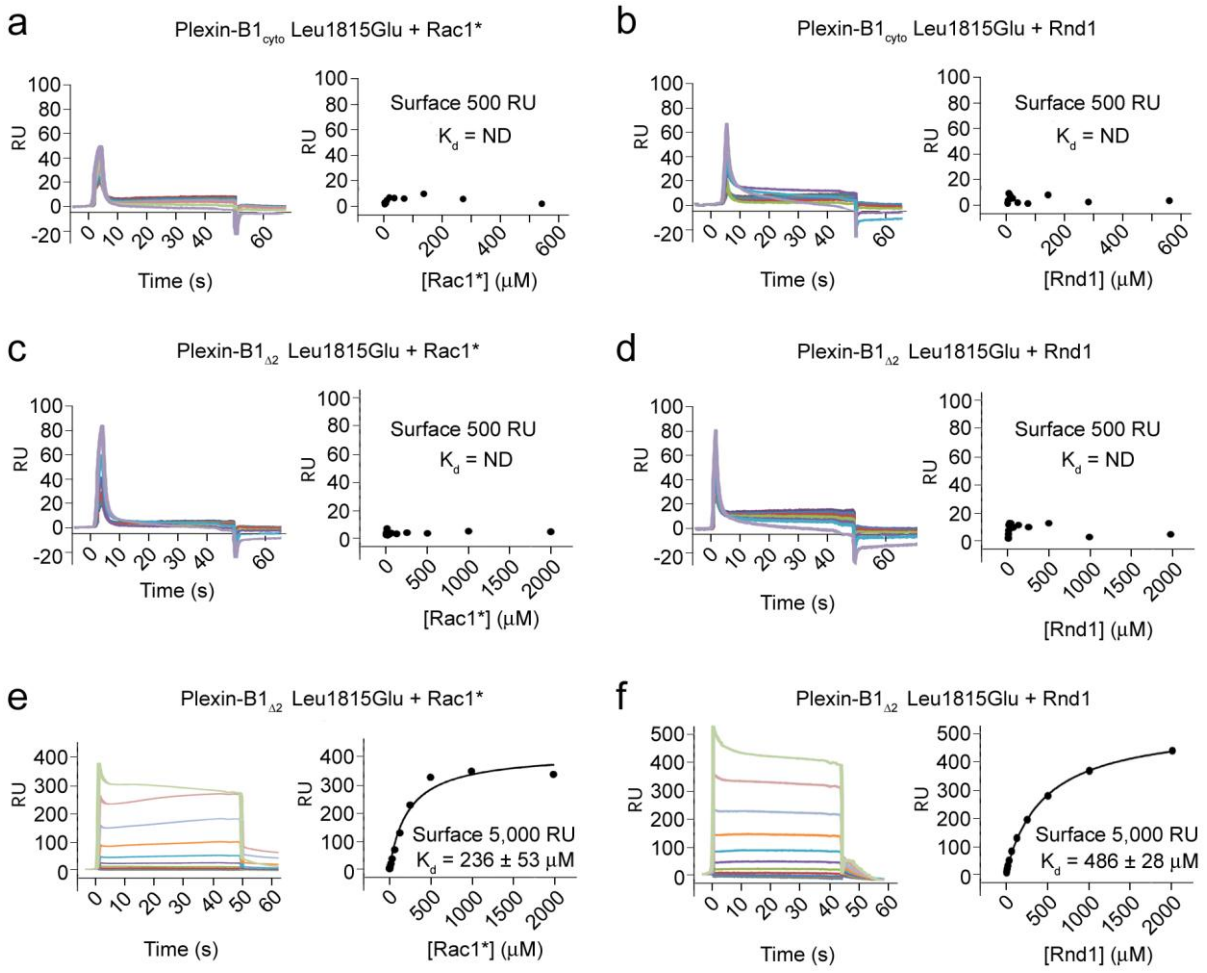


Figure 5.20: Binding of Rac1* and Rnd1 to Plexin-B1$_{cyto}$ and Plexin-B1$_{\Delta 2}$. Data are presented as in Figure 5.10. (a) Plexin-B1$_{cyto}$ Leu1815Glu + Rac1*, 500 RU loaded on the chip, (b) Plexin-B1$_{cyto}$ Leu1815Glu + Rnd1, 500 RU loaded on the chip, (c) Plexin-B1$_{\Delta 2}$ Leu1815Glu + Rac1*, 500 RU loaded on the chip, (d) Plexin-B1$_{\Delta 2}$ Leu1815Glu + Rnd1, 500 RU loaded on the chip, (e) Plexin-B1$_{\Delta 2}$ Leu1815Glu + Rac1*, 5,000 RU loaded on the chip, (f) Plexin-B1$_{\Delta 2}$ Leu1815Glu + Rnd1, 5,000 RU loaded on the chip. Data for binding of Rac1* or Rnd1 to Plexin-B1$_{cyto}$ Leu1815Glu with 5,000 RU loaded on the chip can be found in Figure 5.10d and h, respectively.

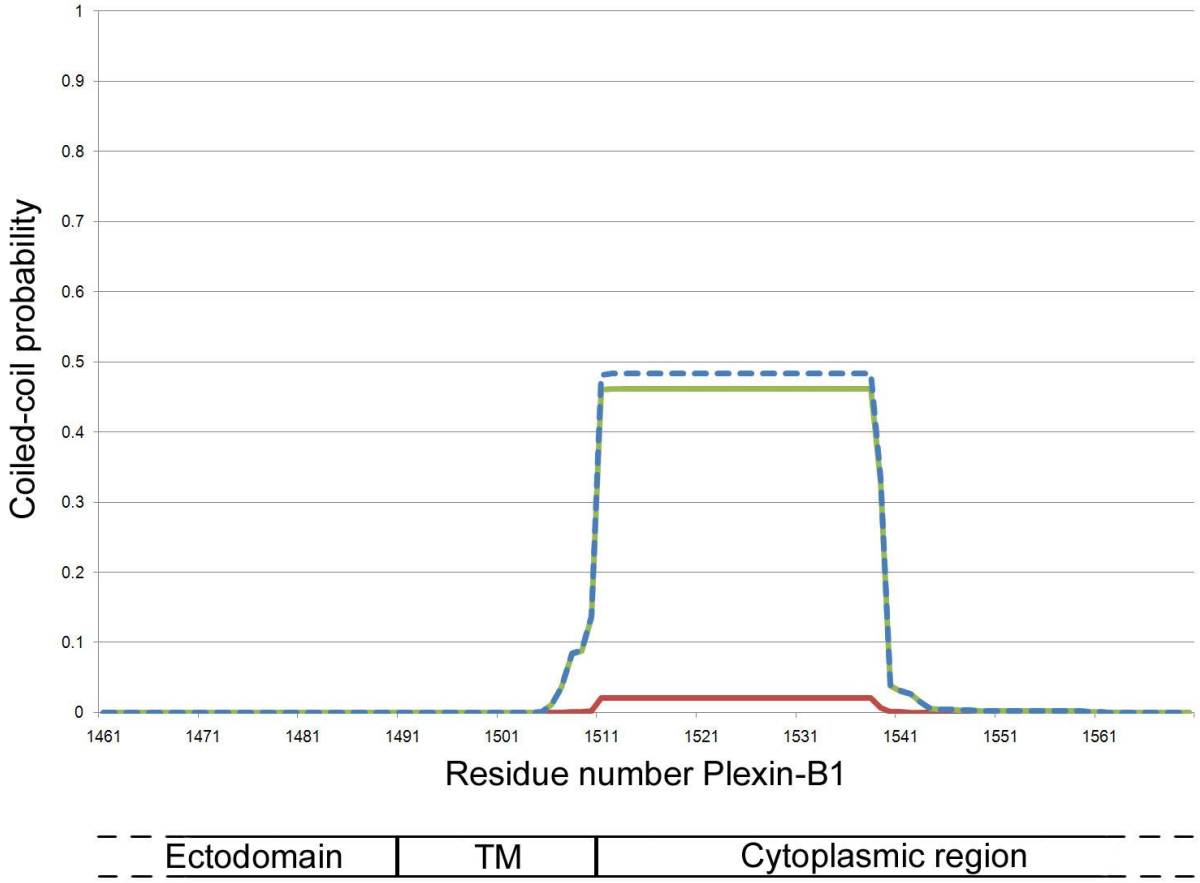


Figure 5.21: The juxtamembrane helix is predicted to form a trimeric coiled-coil. Coiled-coil probabilities were calculated for human Plexin-B1 using MultiCoil (http://groups.csail.mit.edu/cb/multicoil/cgi-bin/multicoil.cgi) and plotted against residue number. Overall probabilities are shown in dashed blue, trimeric coiled-coil probabilities in green and dimeric coiled-coil probabilities in red. The domain organization corresponding to the residue numbers is shown under the graph.

5.8 Discussion

Plexin-semaphorin signalling is dependent on signals from both the extra- and intracellular side. Several studies have shown that both semaphorin binding on the outside and RhoGTPase binding on the inside of the cell are required for plexin activity to occur [49,197]. The nature of these signals and how they are integrated into a single output, namely

RasGAP activity, has been a critical question in this field and several models have been proposed [35-37,43,44,208].

Recently, several structures of truncated plexin ectodomains in complex with their cognate semaphorins have been reported [35-37]. Despite ranging across three different classes, all of these ectodomain complexes share the same overall architecture with one semaphorin dimer bringing together two plexin monomers. In combination with a detailed biophysical and cellular characterization, these structural data have led to the proposal that the bivalency effect is a prerequisite for plexin signalling [35,37].

For the cytoplasmic region, our structures of Plexin-B1 in complex with Rac1* do not show major structural rearrangements when compared to the apo-structure of Plexin-B1 [44]. For the one-to-one complex, Rac1* is positioned distant from the Ras binding site on the Plexin-$B1_{\Delta1}$ molecule. This excludes the possibility of a direct interaction or regulation of RasGAP activity by the RhoGTPase. Instead, the three-fold complex reveals an additional binding site on a neighbouring Plexin-$B1_{cyto}$ molecule that is in close proximity to the predicted Ras binding site. This interaction may result in a small conformational change in the Ras binding region, although a detailed analysis of these changes cannot be made due to the low resolution of our data. It is, however, noteworthy that allosteric regulation of Ras binding by RhoGTPase binding has been proposed by He et al. based on a homology model of the Plexin-A3-Rnd1 complex [43].

The occurrence of a three-fold arrangement in crystals of Plexin-B1-Rac1 complexes appears to be dependent on the juxtamembrane, N-terminal helix. Physiologically, this region connects the intracellular domain with the transmembrane and extracellular region. This

suggests a mechanism by which both semaphorin binding on the outside and RhoGTPase binding on the inside are connected to result in RasGAP activity (Fig. 5.22 and Fig. 5.23). The first step in this model is binding of the RhoGTPase to binding site A of the intracellular domain. Although RhoGTPase binding has been shown to be a prerequisite for Ras binding it is not sufficient to trigger signalling [49]. Semaphorin binding on the outside of the cell may result in clustering of the receptors [35] either from an autoinhibited, monomeric or dimeric state [43,44,208]. Such extracellular rearrangement could be transmitted to the intracellular N-terminal helix. The rearrangement of this juxtamembrane helix would free up binding site B, allowing the RhoGTPase to bridge two plexin molecules and stabilize the three-fold arrangement. Formation of a trimeric cluster could result in the proper positioning of the catalytic machinery allowing RasGAP activity to occur, since it has been shown that clustering of the intracellular domain is crucial for this activity [49]. In summary, the data presented here suggests that receptor clusters nucleated by the dimeric complex on the extracellular side and the trimeric complex on the intracellular side will integrate both RhoGTPase and Semaphorin binding into a single signalling output.

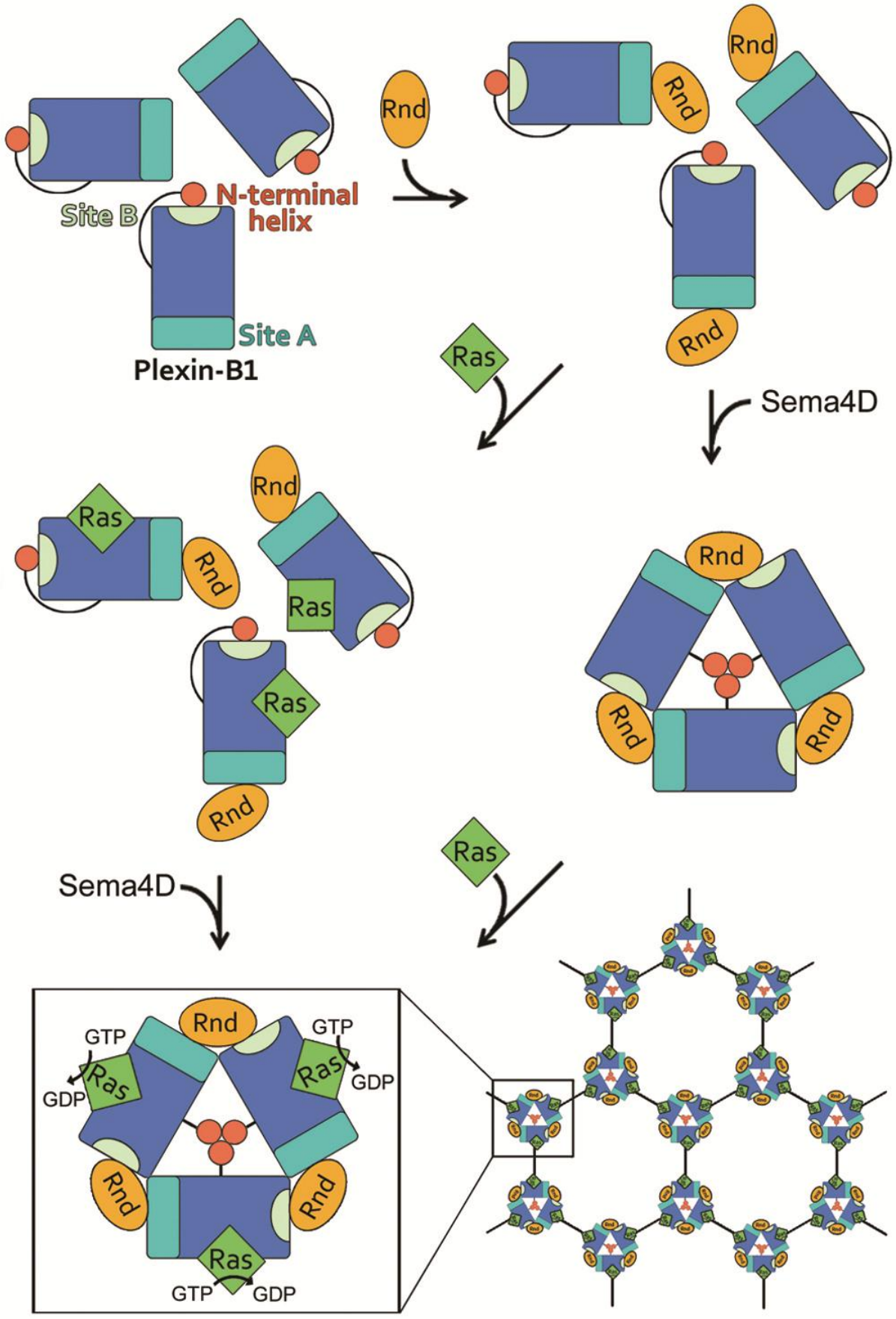

Site B
N-terminal helix
Site A
Plexin-B1
Rnd
Ras
Sema4D
GTP
GDP

Fig. 5.22: Model for intracellular plexin signalling. The Ras-GAP activity of Plexin-B1 is a result of a two-step signalling process. Step one would consist of Rnd1 binding to the intracellular Plexin-B1 RBD region. Binding of Ras molecules to this complex cannot result in GTPase activity unless a second binding event, involving extracellular semaphorin-mediated plexin clustering, induces formation of the intracellular three-fold Plexin-B1-Rnd1 complexes. Such an arrangement is stabilised by the interaction of Rnd1 with a novel binding site on a neighbouring Plexin-B1 molecule, exposed following the displacement of a juxtamembrane helix at the N-terminal of the plexin intracellular region. In this clustered arrangement RasGAP activity can occur and Ras gets inactivated. The trimer on the inside, in conjunction with semaphorin-plexin dimers on the extracellular side, may lead to the formation of an extended signalling array. R-Ras/M-Ras (Ras) and Rnd1 (Rnd) are depicted in their GTP-bound form. Plexin-$B1_{cyto}$ is depicted in blue rectangle, with sites A and B highlighted. The mobile juxtamembrane helix at the N-terminal of the plexin intracellular region is indicated by a red disc. The two GTPases are shown as orange ovals (Rnd1) and green diamonds (Ras), respectively. Semaphorin induced dimerization of the plexin ectodomain is indicated by black lines.

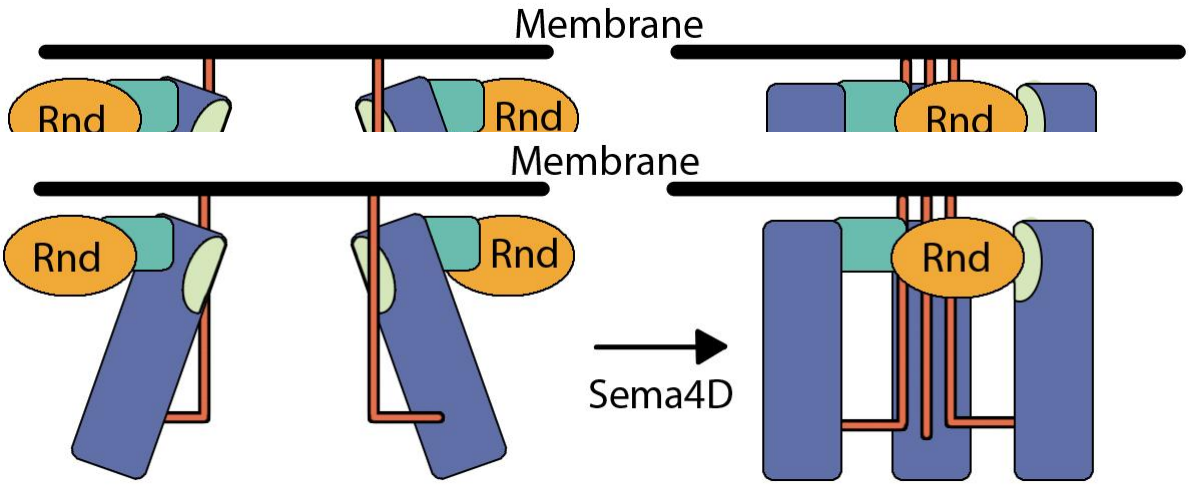


Fig. 5.23: Schematic view of intracellular plexin signalling along the z-axis. Only the intracellular regions are shown for clarity. Colour coding as in Fig. 5.22. The left panel shows the Plexin-B1-Rnd complex with the N-terminal helix blocking site B. The right panel shows the trimeric conformation after Sema4D binding with the N-terminal helices likely forming a trimeric coiled-coil, thus exposing site B and allowing its interaction with Rnd. In the right panel the spectator is looking onto one of the corners of the triangle, thus two of the three Rnd molecules are hidden.

Chapter 6

Structural studies of the Repulsive Guidance Molecule B in complex with its receptor Neogenin

Chapter 6 Structural studies of the Repulsive Guidance Molecule B in complex with its receptor Neogenin

6.1 Summary

Repulsive guidance molecules (Rgms) form a recently discovered protein family that has functions stretching from repulsive axonal guidance, inhibition of leukocyte migration and regulation of the iron master-regulator hepcidin. They are glycosylphosphatidylinositol (GPI)-anchored, membrane associated glycoproteins that have crucial roles in two different signalling pathways, the Neogenin and the bone morphogenetic protein (BMP) pathway. Despite a number of biochemical and functional studies nothing is known about the structure of Rgms since they do not show homology to any known protein family. Here the structure of RgmB in complex with its receptor Neogenin is presented. The structure reveals a novel protein fold for Rgms, allows a detailed description of the interface between RgmB and its receptor and shows the importance of a post-translational cleavage event for folding and function. The structural data together with an extensive biochemical and biophysical characterization suggest an explanation of previously identified Rgm disease mutants and a model for Rgm-mediated activation of the Neogenin signalling pathway.

6.2 Introduction

The Rgm family consists of three members, RgmA, RgmB (also called DRAGON) and RgmC (also called Hemojuvelin) [71,82]. RgmA was the first family member to be identified as

repulsive guidance cue for temporal retinal axons [70] and further studies showed that it is highly expressed in the developing and adult central nervous system (CNS) [75]. RgmB has a similar expression pattern, however, its expression does not overlap with that of RgmA [75]. RgmA has been reported to inhibit axonal growth and recovery after spinal cord injury and application of an anti-RgmA antibody counter-acted this effect [101].

In addition to the CNS, RgmA and RgmB can be found in a variety of other tissues including heart, lung and liver [81,209]. Recently, RgmA was found as an inhibitor of leukocyte migration during acute inflammation and thus as guidance cue for leukocyte chemotaxis [210]. Moreover, Muramatsu et al. reported that RgmA is a T-cell activator and that an RGMa-specific antibody effectively suppresses T cell responses and attenuates the clinical course of experimental autoimmune encephalomyelitis therefore making RgmA a promising target to treat multiple sclerosis [211]. RgmB has been shown to be a regulator of immune responses by negatively regulating IL-6 expression in a ligand dependent manner [85].

The receptor for Rgms in this context is Neogenin [72]. Neogenin is a type I-transmembrane protein whose extracellular domain consists of four immunoglobulin (Ig) domains followed by six fibronectin type III (FNIII) domains. It can mediate both chemoattractive and chemorepulsive axon guidance cues by interacting with Rgms or Netrins, respectively [75,95]. Neogenin is closely related to the Netrin receptor *deleted in colorectal cancer* (DCC), however no binding of Rgm to DCC has been observed [104]. The Netrin binding site on DCC has been mapped to the fourth and fifth FNIII domain and is likely to be similar for Neogenin since the two receptors share 67 % homology in these domains and also show similar affinity for Netrin [212]. The Rgm-binding site has been mapped for RgmC to reside on the

fifth and sixth FNIII domain, therefore partly overlapping with the Netrin binding site [74]. So far it is not clear if Rgm and Netrin binding can be accommodated at the same time or if their binding events are mutually exclusive.

In addition to the Neogenin pathway, Rgms also act as co-receptors in BMP signalling [82]. BMPs represent a large family of Transforming Growth Factor-beta (TGF-β)-like ligands [107,213]. They play crucial roles in embryonic development as evidenced by the fact that a functional loss of BMPs or their pathway members are embryonic lethal. Postnatally, BMPs have been shown to regulate hepcidin expression, a master regulator for iron homeostasis [214]. The iron disorder juvenile hemachromatosis, characterized by excess absorption and storage of iron in tissues, has been directly linked to mutations in the co-receptor for BMP, RgmC [215]. These mutations impair the ability of RgmC to increase BMP signalling and thus hepcidin expression in hepatocytes [216]. Recent studies suggest that Neogenin may also be needed for the RgmC-induced increased hepcidin expression and indeed the Rgm-mediated regulation of BMP signalling pathways altogether [111,217].

Moreover, a receptor-independent function for RgmB has been proposed since adhesion of RgmB-expressing dorsal root ganglions to HEK293 cells transfected with RgmB-cDNA was increased, thus suggesting a potential role of RgmB in cell-cell adhesion via homophilic interactions [83].

Rgm family members have a conserved acid-labile Asp-Pro cleavage site which results in non-continuous amino acid chain that is held together by disulfide bonds. RgmC has an additional proteolytic site that can be cleaved by furin-like proprotein convertase [218]. This

then releases the GPI-anchored, membrane-associated glycoprotein and allows it to interact not just with local but also distant receptors.

The aim of this project was to elucidate the structure of the Rgm-family member RgmB. The interactions with its receptor Neogenin were studied both crystallographically and biophysically by means of SPR binding experiments allowing insights into the role of Rgm in Neogenin- as well as BMP-signalling.

6.3 Expression and purification of RgmB and Neogenin

Several constructs of mouse Neogenin were designed of which three were chosen for further study. These constructs included the full-length extracellular domain with small truncations at the N-as well as C-terminus (Neo_{ecto} : residues 63-1082), the fifth and sixth FNIII domains plus additional, membrane-proximal 54 residues (Neo_{56m} : residues 884-1136) and just the fifth and sixth FNIII domains (Neo_{56} : residues 884-1082).

Small scale expression trials in HEK 293T cells showed soluble, secreted expression for all three constructs (see section 2.2.3). All constructs included predicted N-linked glycosylation sites (eight for Neo_{ecto}, one at residue 940 for Neo56m and Neo56; using the NetNGlyc 1.0 server, CBS, Denmark). Thus, large scale expression in HEK293T cells was carried out using the glycosylation inhibitor kifunensine (see section 2.2.3). All proteins were purified using a two-step strategy as described in section 2.2.3. After deglycosylation, analysis of the three Neogenin constructs on SDS-PAGE showed sharp bands for Neo_{ecto} and Neo_{56} indicating successful removal of the sugar moieties (Fig. 6.1) while Neo_{56m} remained rather diffuse and ran at a higher molecular weight than expected from the sequence. Careful inspection of

the sequence revealed that the membrane-proximal region is rich in Gly, Ser and Thr suggesting O-link glycosylation at some of the Ser or Thr residues. Indeed, prediction of O-link sugars using the NetOGlyc 3.1 server (CBS, Denmark) showed at least nine sites for O-link glycosylation.

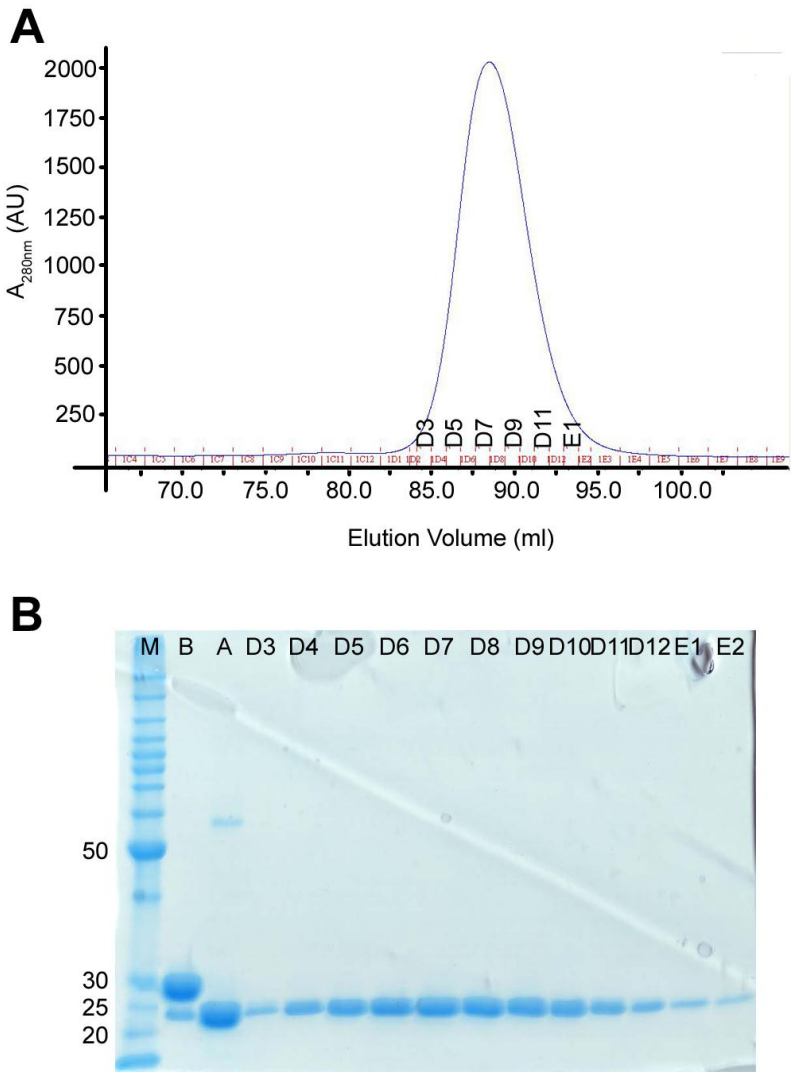


Figure 6.1: Purification of Neo_{56}. A) Gelfiltration profile of an S200 16/60 column after deglycosylation. B) SDS-PAGE analysis of the gelfiltration run shown in A). M: Marker, B: before deglycosylation, A: after deglycosylation. The molecular weight of chosen marker bands is shown in kDa.

All members of the human Rgm family, RgmA, RgmB and RgmC were cloned into the pHLSec expression vector leaving out the GPI anchor and starting at the predicted signalpeptide cleavage site (using the SignalP 3.0 Server, CBS, Denmark). Small scale expression tests showed weak soluble expression for RgmA and strong soluble expression for RgmB. RgmC could not be detected on the anti-His Western-Blot. Since it was known that RgmC can be processed by proprotein convertase to cleave off the C-terminal 70 residues [90], the recognition site for this protease (residues 326-335) was modified by replacing the Arg residues with Ala (Arg326Ala, Arg329Ala, Arg332Ala). This RgmC construct, hereafter called $RgmC_{AAA}$, could be solubly secreted, detected on a Western-Blot and was used for all subsequent studies.

Since RgmB showed the highest levels of expression, this construct was chosen for structural studies and complex formation with Neogenin. Initial large scale expression experiments showed two distinct peaks on a gelfiltration run, corresponding to dimeric and monomeric RgmB (Fig. 6.2). Careful inspection of the signal peptide suggested that the reason for this could be a single Cys residue in close proximity to the predicted cleavage site. This single Cys is not conserved in other Rgm-family members and thus the construct was redesigned to avoid this residue and thus started at residue 53. The redesigned construct showed a single peak, corresponding to a monomeric species and its expression level was approximately 7-fold higher than for the initial construct (Fig. 6.2). All binding studies and mutants were generated using this construct. Crystallisation was ongoing during the redesign of the construct, thus all crystallisation trials were set-up with the initial construct.

Complex formation of RgmB and Neo_{56} for crystallisation was achieved by mixing the glycosylated proteins in a molar ratio of 1 : 1.1 , deglycosylation over night followed by a final purification step on a gelfiltration column. Quantitative complex formation could be achieved and verified by SDS-PAGE.

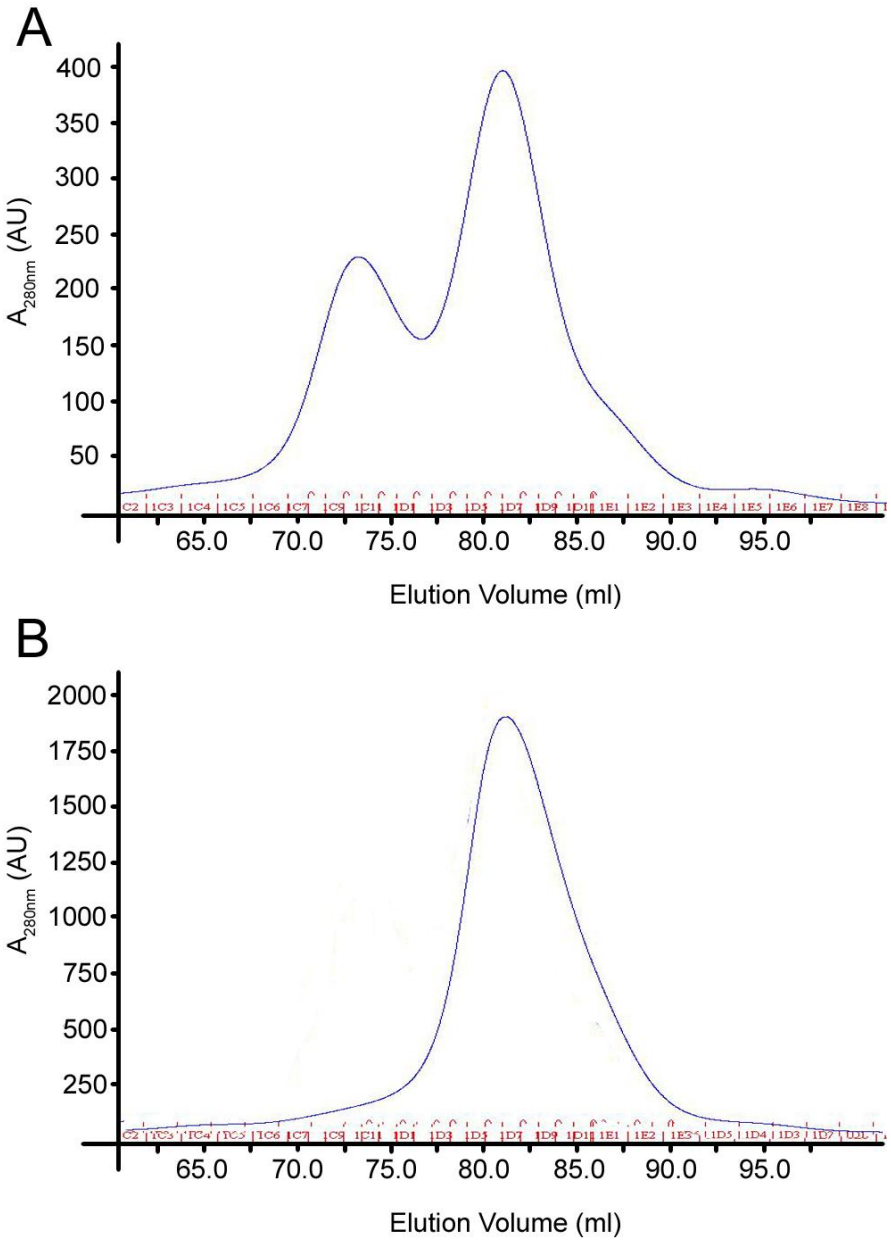


Figure 6.3: Optimization of the RgmB expression construct. A) Gelfiltration profile from an S200 16/60 column of the initial RgmB construct. B) Gelfiltration profile from an S200 16/60 column of the redesigned, optimized construct.

6.4 Crystallization and structure determination of the Neo_{56}-RgmB complex

Several crystal trials were set-up with RgmB and $RgmC_{AAA}$ on their own. Despite screening almost 1000 conditions each, no crystals could be obtained. RgmB was subsequently complexed with Neo_{56} with the intention of stabilizing the possibly flexible RgmB by interaction with its high-affinity receptor. Crystals of Neo_{56} had previously been obtained by the Bjorkman group at CalTech as well as in Strubi (Christian Siebold, personal communication) and the structure was published recently [219]. The Neo_{56}-RgmB complex crystallised initially in 100 mM citric acid, pH 3.5 and 25 % PEG 3350. Several rounds of optimization, in which the precipitant concentration, the pH, the protein to reservoir drop ratios and a variety of different additives were explored, finally lead to crystals resembling bipyramidal tetragons and a complete dataset was collected at beamline ID23-2 at the European Synchrotron Radiation Facility, France, following a helical data collection strategy to ensure both a high-resolution and completeness of the dataset. The complex structure was solved by molecular replacement in PHASER [122] using the structure of the fifth and sixth FNIII domain of human Neogenin (PDB ID: 3P4L [219]) as a search model. Additional electron density for RgmB was immediately discernible. After density modification with Parrot as implemented in CCP4i [120], the initial RgmB polypeptide chain was traced using Buccaneer [120]. Iterative rounds of refinement in autoBUSTER [125] and chain tracing in Buccaneer resulted in a well-defined model for human RgmB that includes residues 134-338 (Fig. 6.3). The N- and C-termini of RgmB could not be traced due to missing electron density and were thus not included in the final model. In addition, the loop region of RgmB between residues 143 and 157 was also disordered and not included in the final model. Neo_{56} has a single

glycosylation site for which discernible electron density was found for the first N-Acetylglucosamine (GlcNAc) sugar residue. Data collection and refinement statistics can be found in Table 5.

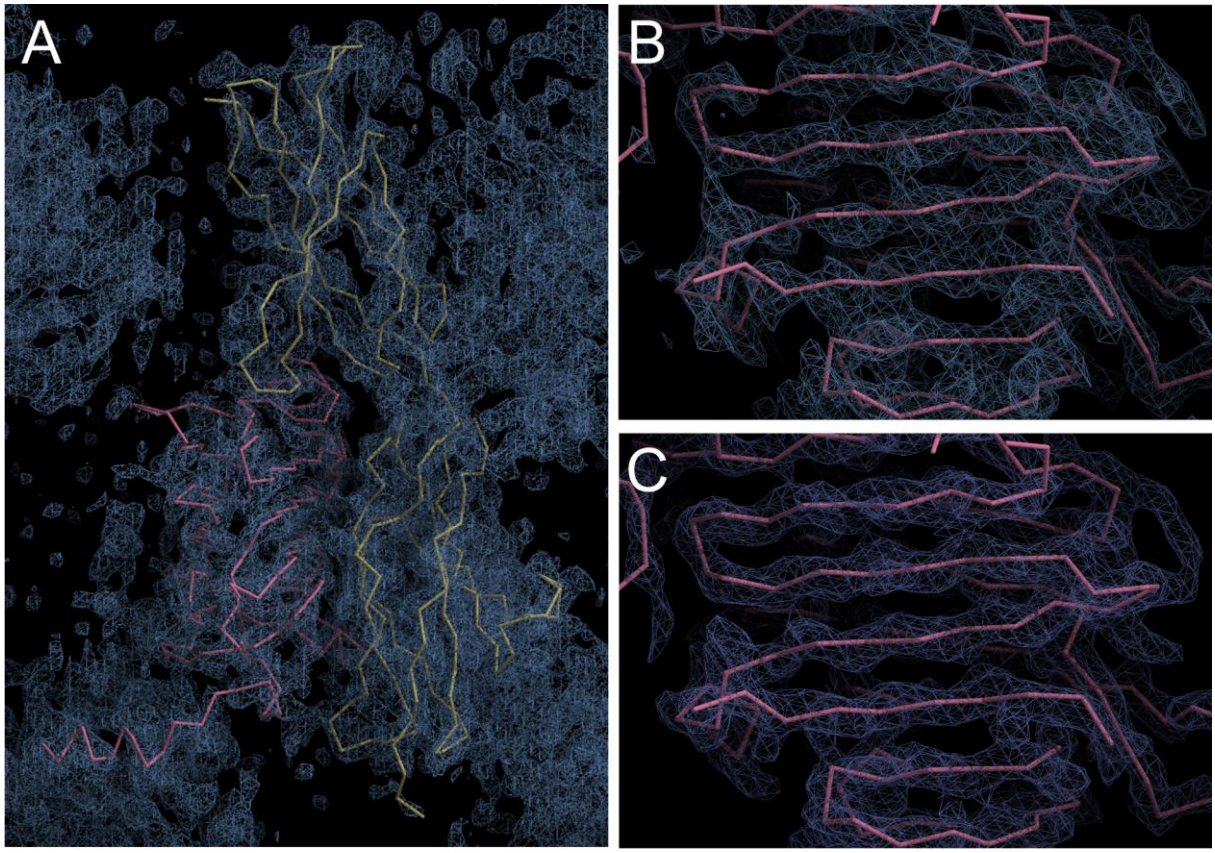


Figure 6.3: Electron density of the Neo_{56}-RgmB complex. (A) Initial electron density map contoure at 1.4 σ after molecular replacement using PHASER [122] and density modification using Parrot [120]. The final model of the Neo_{56}-RgmB complex is represented in lines with RgmB in pink and Neo_{56} in yellow. The orientation is similar to Fig. 6.11. (B,C) Detailed view onto one side of the RgmB beta-sheet. (B) Initial electron density map contoured at 1.4 σ after molecular replacement using PHASER [122] and density modification using Parrot [120]. (C) $2F_O$-F_C map contoured at 1.4 σ of the final refined model.

Stereochemical properties were assessed by MOLPROBITY [127]. Ramachandran statistics are as follows (favored/disallowed (%)): 93.99/0.0. Superpositions were calculated using SHP [204]. Buried surface areas of protein-protein interactions were calculated using the PISA webserver (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html).

Table 5. Data collection and refinement statistics

	Neo_{56}-RgmB
Data collection	
Resolution (Å)	3.10 (3.15-3.10)
Space group	$P4_12_12$
Cell dimension	
a, b, c (Å)	118.9, 118.9, 94.4
α, β, γ (°)	90, 90, 90
Redundancy	12.8 (13.1)
Completeness (%)	100.0 (100.0)
R_{merge} (%)	17.0 (99.8)
Avg I/σ	15.9 (2.6)
Refinement	
Resolution	30.00 - 3.10
No. Reflections	12706
R_{work}/R_{free}	20.3 (24.1)/23.5 (31.2)
No. Atoms	
Protein	2958
Sugar	14
B-factors	
Protein	67.2
Water	94.4
r.m.s. Deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.020

[a] Numbers in parenthesis are for the highest resolution shell

[b] R_{free} equals the R_{work} against 5% of the data removed prior to refinement.

6.5 The structure of RgmB reveals a novel protein fold

The structure of RgmB in complex with Neo_{56} was determined to a resolution of 3.1 Å and refined to a crystallographic R-factor of 20.3 % (R_{free} = 23.5 %, Table 3). Until now, no structural information was available for RgmB, thus this section will focus on the RgmB structure and fold alone. In the next section, the complex between RgmB and Neo_{56} will then be described in detail.

The middle domain of RgmB (residues 134-338) forms an eleven stranded beta-sheet with four additional, short helices at the C-terminus (Fig. 6.4). Both the N- and C-termini of the domain are linked to the beta-sheet by three disulfide pairs formed between Cys139-Cys226, Cys162-Cys312, and Cys181-Cys316. The termini of the domain point into opposite directions and into the solvent channels of the crystal, suggesting that the N- and C-terminal domains which could not be traced in the crystal structure are flexible and do not associate with the middle domain. The core of the middle domain forms a tightly aligned beta-sandwich with seven strands on one sheet and four strands on the other. The topology and the order of the beta-strands does not show any homology to previously determined structures when compared using SSM [220], FATCAT [221] or GANGSTA+ [222] (Fig. 6.5). Thus, this domain is likely to represent a novel protein fold.

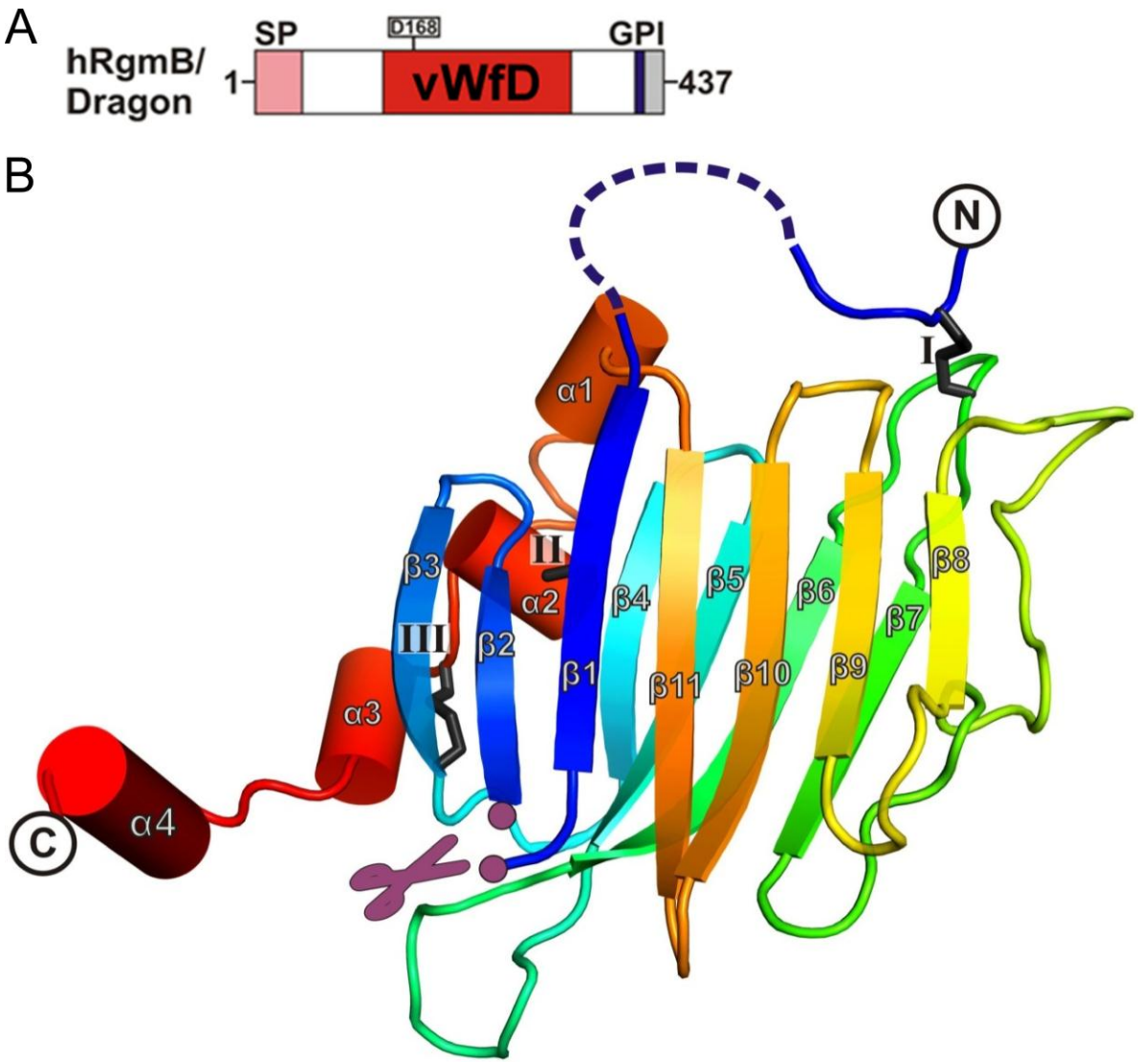


Figure 6.4: Structure of the middle domain of human RgmB. (A) Domain organization of RgmB. The acid labile cleavage site is labelled with D168. The middle domain is abbreviated with vWfd (von Willebrand factor d domain); SP, secretion peptide; GPI, GPI anchor. (B) Structure of the middle domain human RgmB. The protein is coloured in rainbow from N- (blue) to C-terminus (red). Secondary structure elements are labelled. The three disulfide bonds are represented as black sticks and are marked with Roman numerals. The auto-proteolytic cleavage site is labelled in purple and highlighted by a scissor. Residues 143 to 157 could not be traced in the final model and are depicted as dotted line.

The crystal structure revealed that the middle domain is auto-catalytically cleaved between the acid-labile Asp168-Pro169 bond (Fig. 6.6). Asp-Pro bonds have previously been found to

be hydrolysed during exposure to low pH values [223-226]. 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. This autocatalytic, posttranslational modification is likely to occur in the late secretory pathway since the pH in the trans-Golgi network is slightly below 6.0 [227] and can approach pH 5.2 in the secretory granules [228].

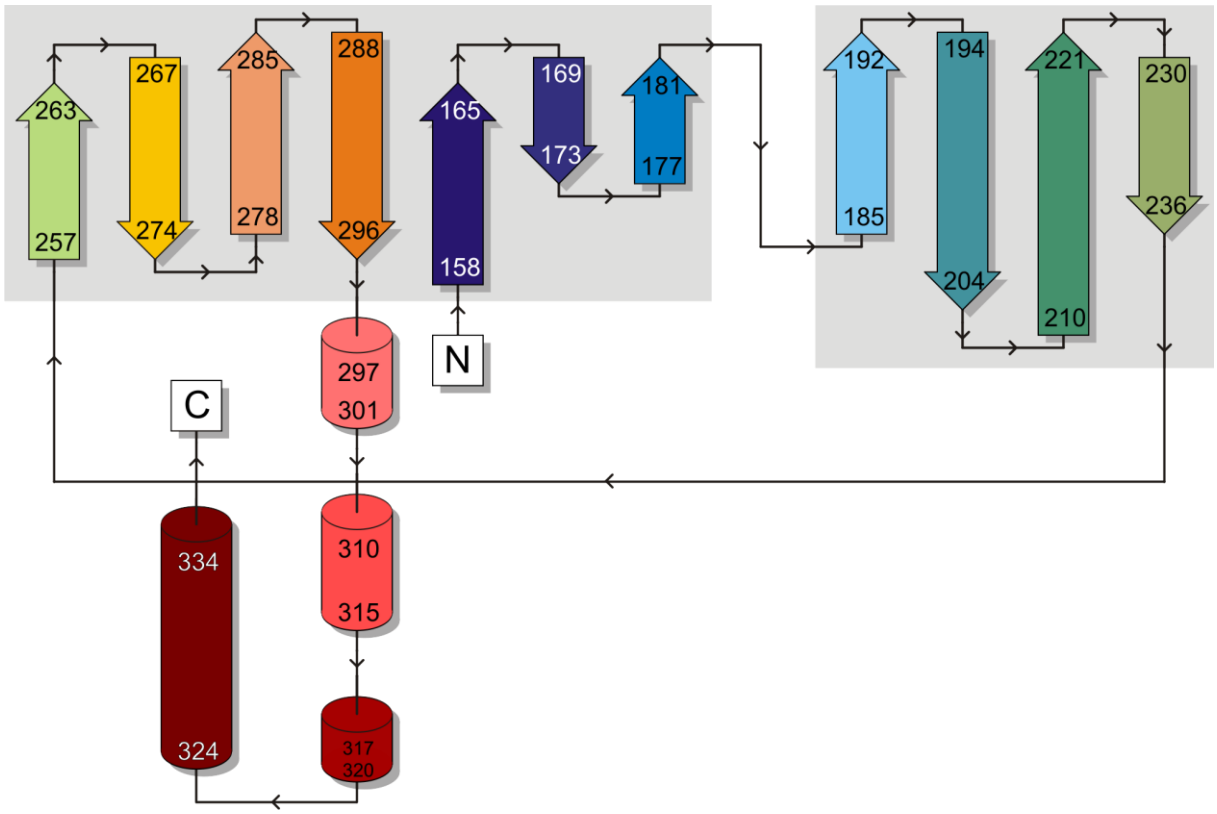


Figure 6.5: Topology diagram of RgmB (adapted from PDBSum [238]). Secondary structure elements are shown with an arrow (β-strand) or cylinder (α-helix) and are coloured as a rainbow with the N-terminus shown in blue and the C-terminus shown in red.

As a result of the auto-catalytic cleavage the sidechain of Pro169 is deeply buried in the protein core and thus adds to the structural integrity of the domain (Fig. 6.6). The positively charged N-terminus of Pro169 is stabilized by interaction with the carbonyl oxygen of Cys181 and Thr211 and the sidechain hydroxyl group of Thr211. The C-terminus of Asp168 is oriented towards the solvent.

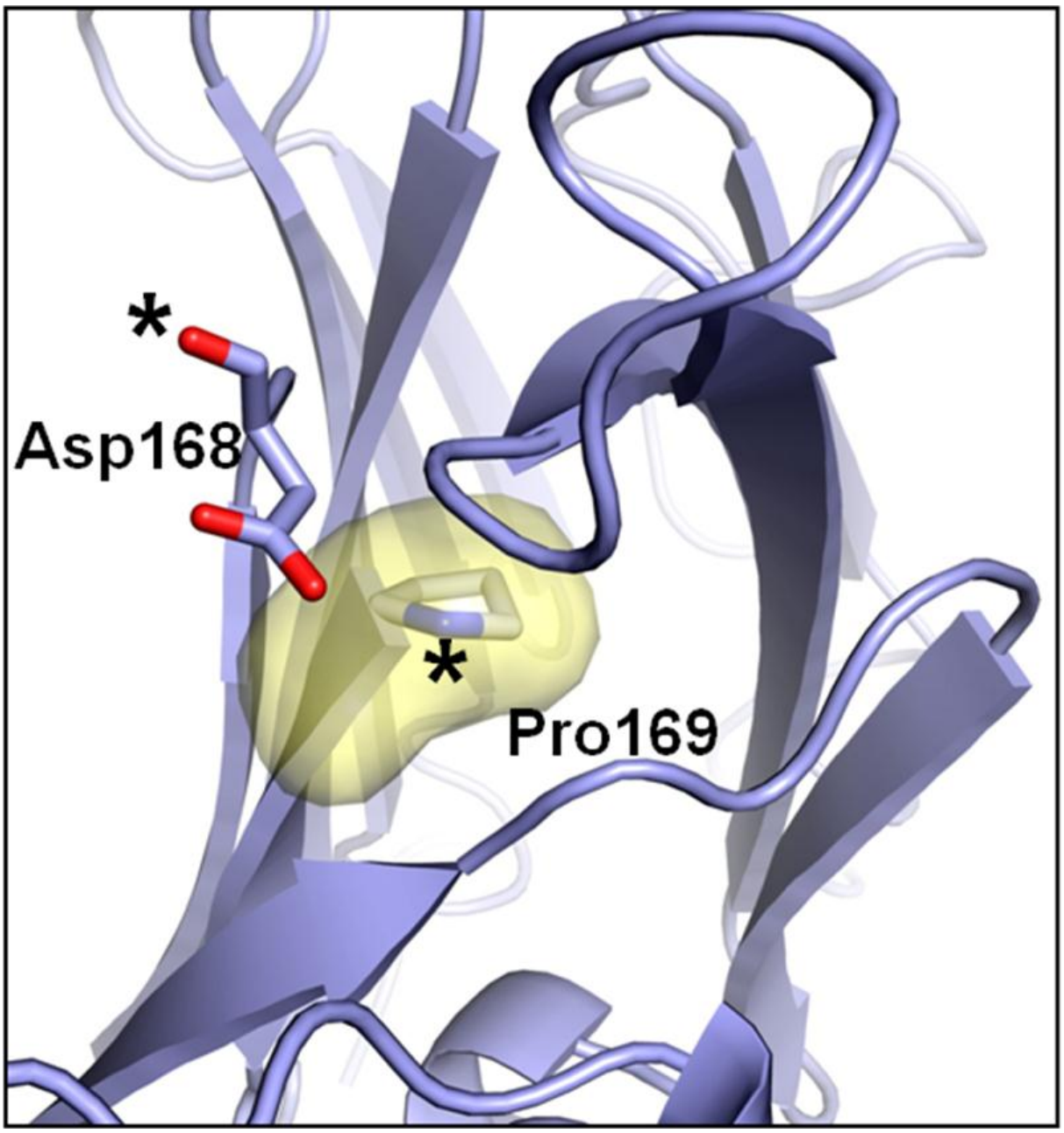


Figure 6.6 Detailed view onto the Asp168-Pro169 cleavage site. Asp168 and Pro169 are represented as sticks. A surface representation of Pro168 is shown in faint yellow. The asterisks mark the N- and C-terminus, respectively.

Since the Pro168 sidechain can only adopt its buried conformation after cleavage of the acid-labile Asp168-Pro169 bond, the functional importance of this modification was further investigated. Indeed, the Asp-Pro dipeptide is conserved amongst different Rgm family members and across different species, from human to zebrafish (Fig. 6.7). In addition, some of the published disease-related mutants of RgmC that lead to juvenile hemachromatosis cluster around the cleavage site (Fig. 6.8). In total there are 14 mutations on RgmC that have been linked to juvenile hemachromatosis so far [215,229,230]. Ten of these map onto the middle domain of RgmB (Fig. 6.8), whereas the remaining four reside on the N-terminal domain. All mutations were cloned into the $RgmC_{AAA}$ background and tested for soluble expression in a small scale expression test (Fig. 6.9). Only two of the mutant proteins, $Gly99Arg^{RgmC}$ and $Leu101Pro^{RgmC}$, could be solubly expressed. All other mutants, including the cleavage site mutants, $Ala168Asp^{RgmC}$, $Phe170Ser^{RgmC}$ and $Asp172Glu^{RgmC}$ were not secreted, thus giving an explanation for their disease-related phenotype. This further strengthens the hypothesis that the auto-catalytic cleavage of the Asp168-Pro169 bond that can be found across all Rgm-family members and species is essential for the structural integrity of the middle domain and indeed the entire protein.

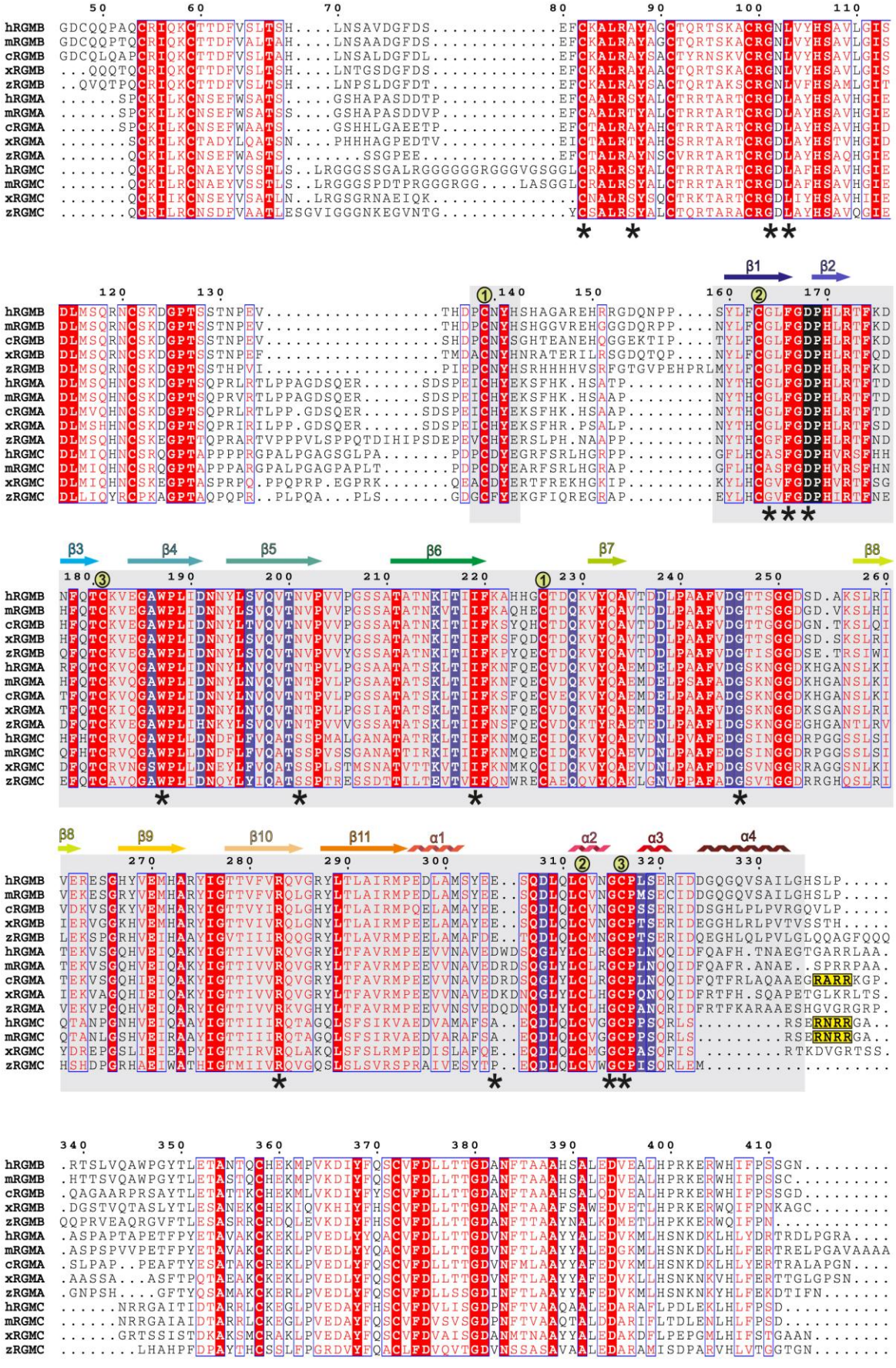

hRGMB
mRGMB
cRGMB
xRGMB
zRGMB
hRGMA
mRGMA
cRGMA
xRGMA
zRGMA
hRGMC
mRGMC
xRGMC
zRGMC
β1
β2
β3
β4
β5
β6
β7
β8
β9
β10
β11
α1
α2
α3
α4

Figure 6.7: Sequence alignment of different Rgm family members from different species. The Rgm sequences were aligned using MULTALIN (bioinfo.genotoul.fr/multalin/multalin.html) and formatted with ESPRIPT (espript.ibcp.fr/ESPript/ESPript/). The signal peptide as well as the C-terminal hydrophobic sequence that is cleaved upon addition of the GPI anchor are not included in the alignment. Secondary structure elements are colour coded following the same pattern as in Fig. 6.4. Disulfide bonds are marked by yellow circles and numbered accordingly. The Asp-Pro cleavage site is highlighted in black. RgmC mutations causing juvenile hemachromatosis are marked with an asterisk. Residues involved in the interaction with Neo_{56}, as described in Chapter 6.6, are highlighted in blue. The furin-like proprotein convertase cleavage site is highlighted with a yellow box.

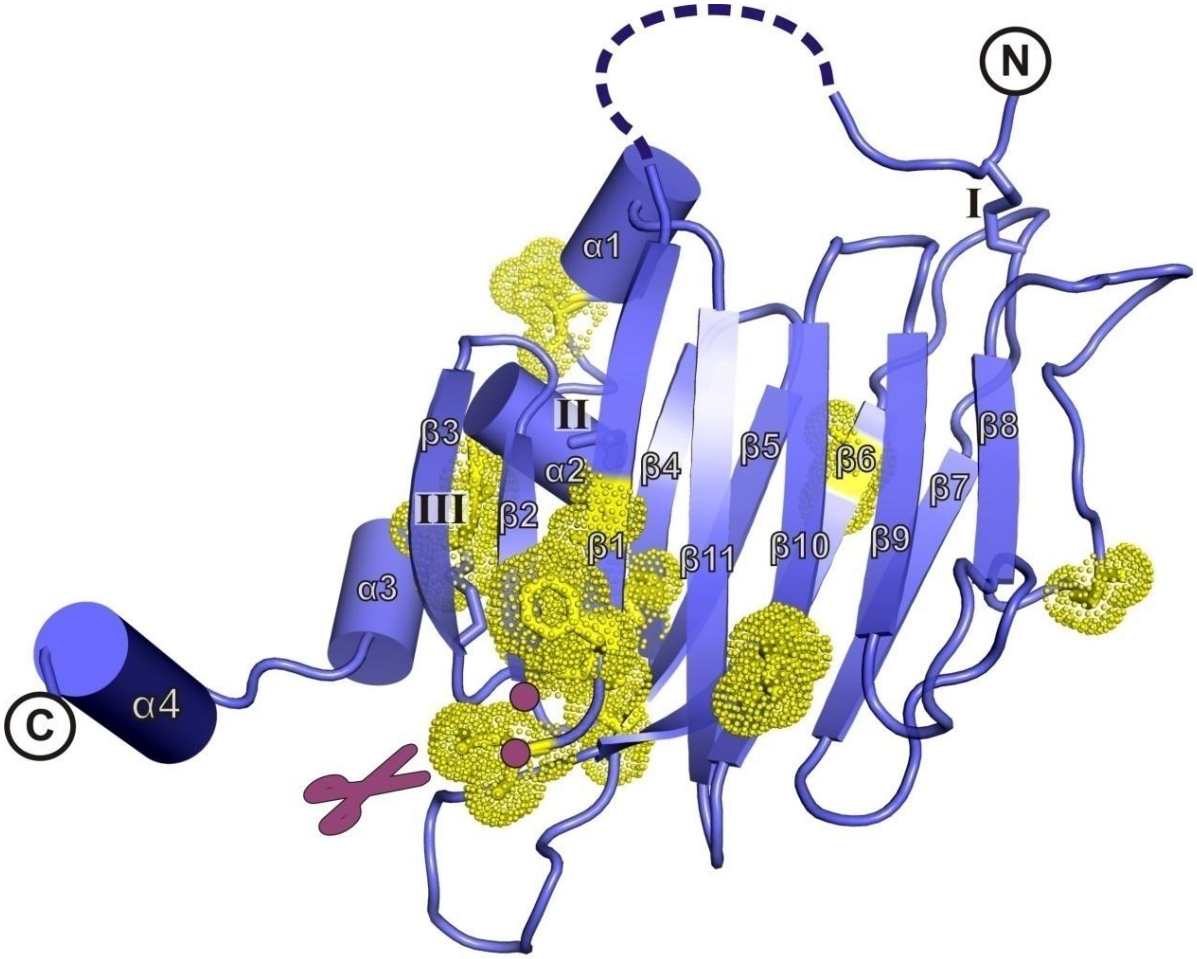


Figure 6.8: RgmC mutations causing juvenile hemachromatosis mapped onto the RgmB structure. Ten of the reported mutations map onto the middle domain of RgmC and equivalent residues on RgmB have been highlighted as yellow sticks with and space filling spheres.

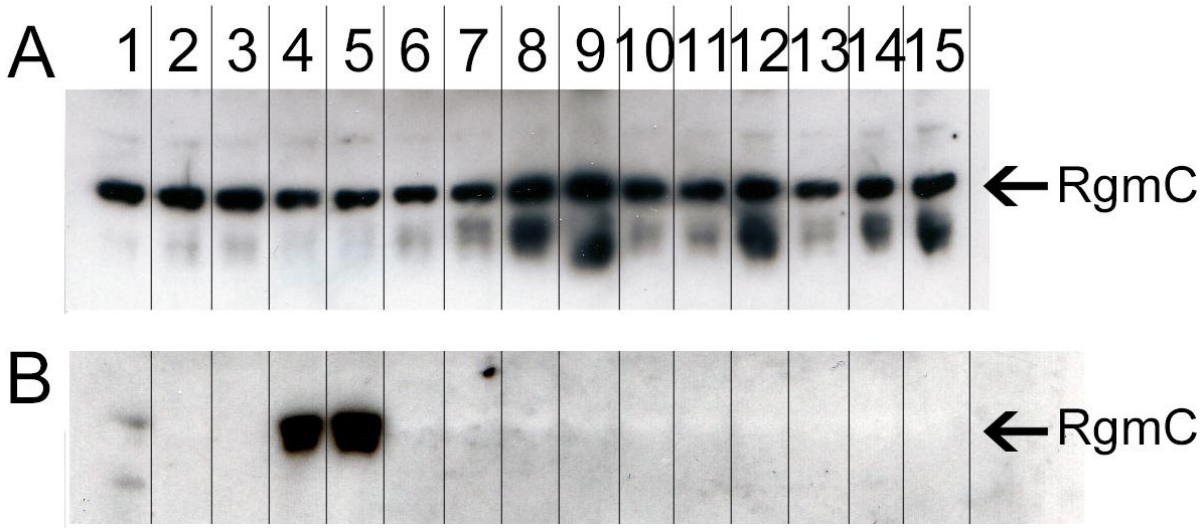


Figure 6.9: Anti-His$_6$ Western Blot of the small scale expression test of the disease-related RgmC mutants. All mutants were generated in the RgmC$_{AAA}$ background. (A) Cells; (B) Supernatant. 1: wild-type, 2: Cys80Arg, 3: Ser85Pro, 4: Gly99Arg, 5: Leu101Pro, 6: Ala168Asp, 7: Phe170Ser, 8: Asp172Glu, 9: Trp191Cys, 10: Ser205Arg, 11: Ile222Asn, 12: Gly250Val, 13: Arg288Trp, 14: Gly320Val, 15: Cys321Trp. The wild-type protein in lane 1 does not show a remarkable band due to problems during gel-loading. In subsequent experiments the wild-type RgmC$_{AAA}$ construct could be successfully secreted.

The middle domain of members of the Rgm-family has sequence similarity with a large protein domain family called von Willebrand factor type D (vWfD) domains [70,231]. No structural information is known for this family and there is little sequence conservation amongst the family members (Fig. 6.10). Members of the Rgm family show sequence similarity in the N-terminal region of the domain. In addition, the autocatalytic Asp-Pro cleavage site can also be found in other vWfD representatives. However, Rgm family members are missing the characteristic GlyLeuCysGly motif towards the C-terminus of the domain [231,232]. This suggests that middle domain can be classified as von Willebrand factor type D-like domain.

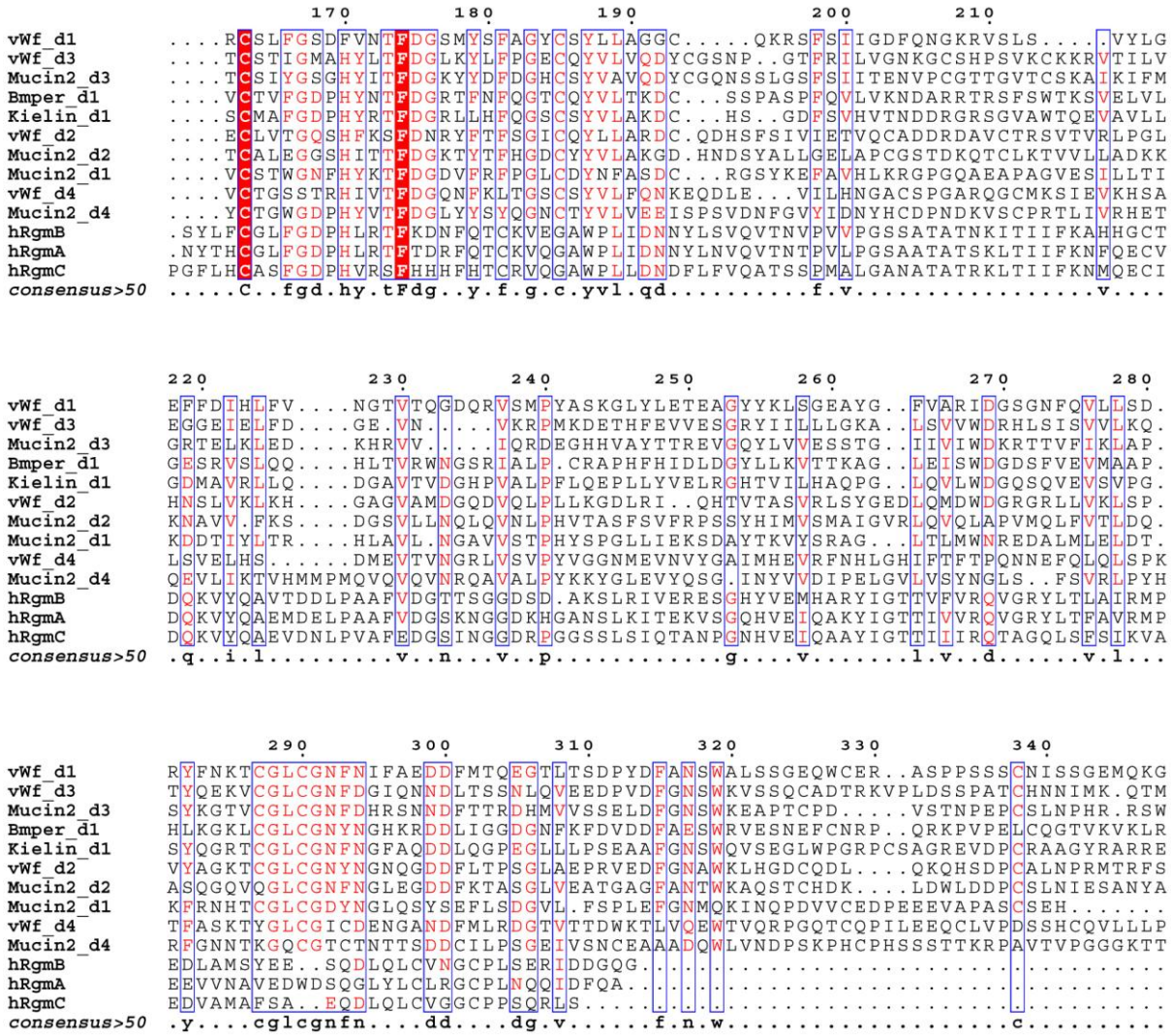


Figure 6.10: Sequence alignment of a selection of vWfD domains. The numbering is in accordance with RgmB. sequences were aligned using MULTALIN (bioinfo.genotoul.fr/multalin/multalin.html) and formatted with ESPRIPT (espript.ibcp.fr/ESPript/ESPript/)

6.6 The interaction between RgmB and its receptor Neogenin

Crystallisation of the Neo_{56}-RgmB complex did not just allow the structural characterization of RgmB and its novel fold but also provided the structural basis for the interaction between RgmB and its receptor. Previous studies have mapped the RgmB binding site on Neogenin onto the fifth (FN5) and sixth FNIII (FN6) domain. Indeed, the complex structure reveals that

RgmB binds exclusively to these two regions (Fig. 6.11). The interface is slightly hydrophobic with 16 additional hydrogen bonds and two salt bridges formed between the two proteins (Table 6). In total, the interface covers a surface area of 1102 $Å^2$. It can be divided into two separate units, one between RgmB and the Neogenin-FN6 and one between RgmB and the L3-loop region of the Neogenin-FN5 domain. The former includes the majority of contacts, covering a total surface area of 814 $Å^2$ or 74 % of the total interface and is hereafter called site A. It is mediated by four strands of the RgmB β-sheet that is facing the Neogenin-FN6 domain with their sheets being orientated orthogonal towards each other (Fig. 6.11). Site A includes 13 of the 16 hydrogen bonds and all salt bridges (Table 6, Fig. 6.12, Fig. 6.13). Thus, it is likely to be the main driving force for complex formation. The interface between RgmB and Neogenin-FN5 is considerably smaller with a surface area of 288 $Å^2$ or 26 % of the total interface area and is hereafter called site B. The interaction includes three hydrogen bonds and is centered between residues 238 and 248 of RgmB and the L3 loop region of Neogenin (Table 6, Fig. 6.14).

The residues involved in the interaction are conserved amongst different Rgm-family members and across different species thus suggesting that the interface observed here reveals a common architecture for Rgm-Neogenin interactions (Fig. 6.7, Fig. 6.15). Accordingly, the interface residues on Neogenin are also conserved across different species (Fig. 6.15, Fig. 6.16).

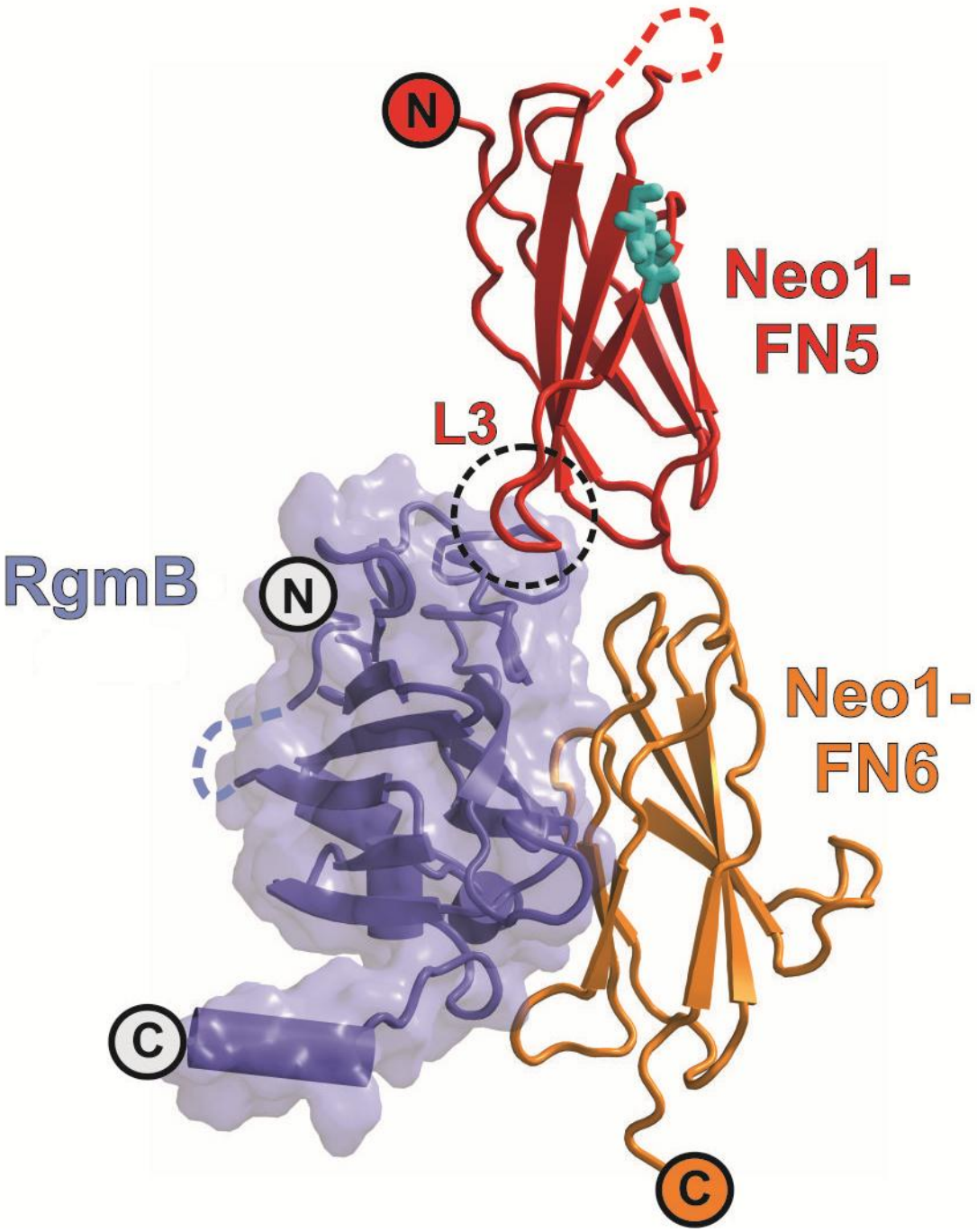


Figure 6.11: The RgmB-Neo56 complex structure. The middle domain of RgmB is shown in blue, the FN5 domain of Neogenin in red and the FN6 domain of Neogenin in orange. The GlcNAc sugar residue on Neogenin-FN5 is shown as sticks and coloured in teal. The N- and C-termini are labelled for both proteins. The L3-loop on FN5 is labelled and highlighted with a dotted circle. Residues 912-916 of Neogenin could not be traced and are represented by a dotted line.

Hydrogen bonds		
RgmB	Distance in Å	Neo_{56}
THR 200[OG1]	3.18	ASP 991[OD2]
ALA 186[N]	3.24	THR 993[OG1]
LEU 318[N]	3.10	GLU 998[OE1]
GLN 198[NE2]	3.11	TRP1008[O]
THR 217[OG1]	2.75	GLN1009[OE1]
SER 196[OG]	3.47	ASN1044[O]
GLN 198[NE2]	2.89	LEU1046[O]
GLY 246[O]	2.78	ASN 929[ND2]
ASP 238[OD1]	3.10	ALA 932[N]
ASP 245[O]	2.90	LEU 956[N]
ALA 186[O]	3.14	THR 993[OG1]
ASP 308[OD2]	3.02	LYS 997[NZ]
ALA 186[O]	3.18	ASN1007[ND2]
THR 200[OG1]	3.11	GLN1009[NE2]
ASP 191[OD2]	2.94	ARG1045[NH2]
ASP 308[OD1]	2.96	GLN1049[NE2]

Salt bridges		
RgmB	Distance in Å	Neo_{56}
ASP 308[OD2]	3.02	LYS 997[NZ]
ASP 191[OD2]	3.51	ARG1045[NH1]
ASP 191[OD1]	3.11	ARG1045[NH2]
ASP 191[OD2]	2.94	ARG1045[NH2]

Table 6: List of all hydrogen bonds and salt bridges in the Neo_{56}-RgmB complex. The structure was analysed using the PISA webserver (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html).

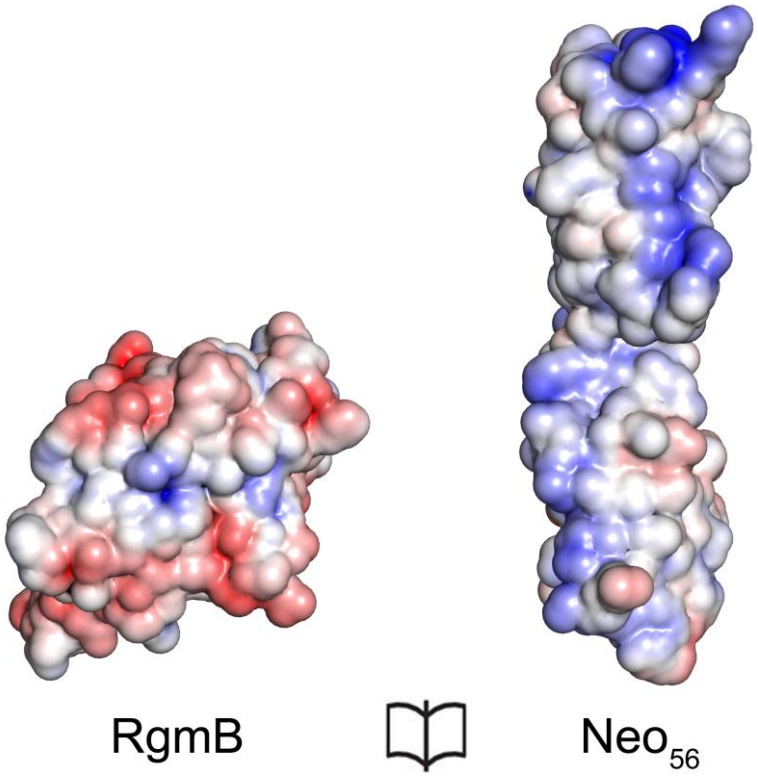


Figure 6.12: Open book representation of the Neo56-RgmB complex. The proteins are represented with their surfaces coloured according to electrostatic potential, calculated in APBS (http://www.poissonboltzmann.org/apbs), and mapped onto the surface with the color code ranging from -10 kT/e (bright red) to +10 kT/e (dark blue).

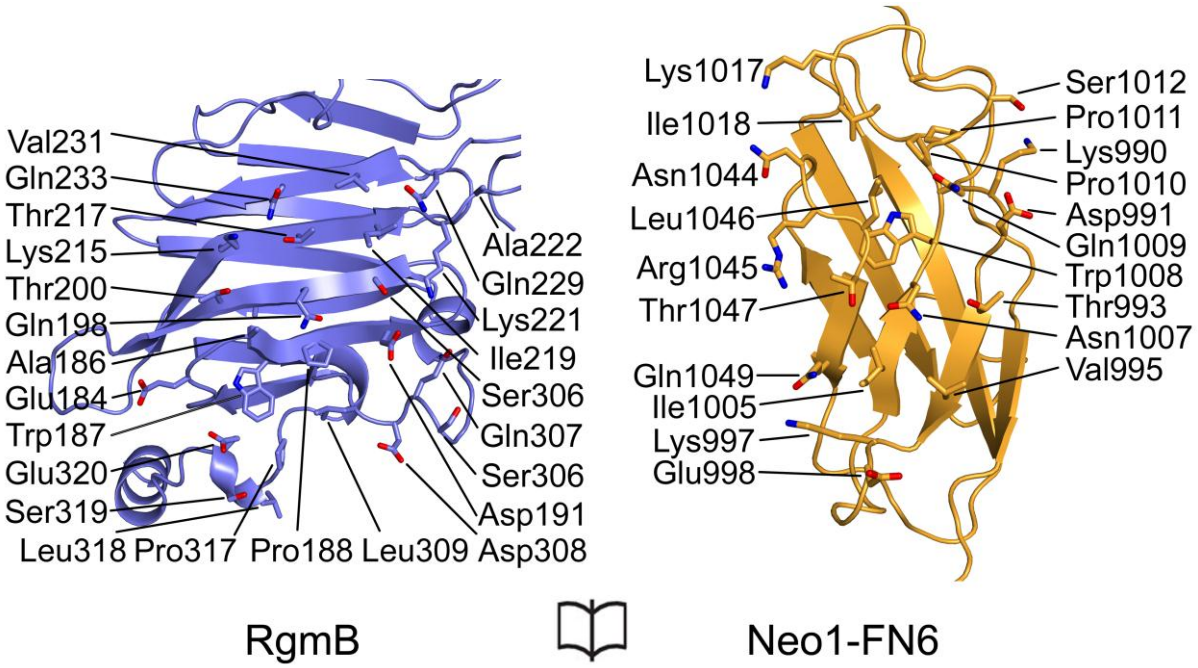


Figure 6.13: Detailed view onto binding site A. Residues involved in the interaction are shown as sticks and labelled accordingly. The colour coding follows Fig. 6.11.

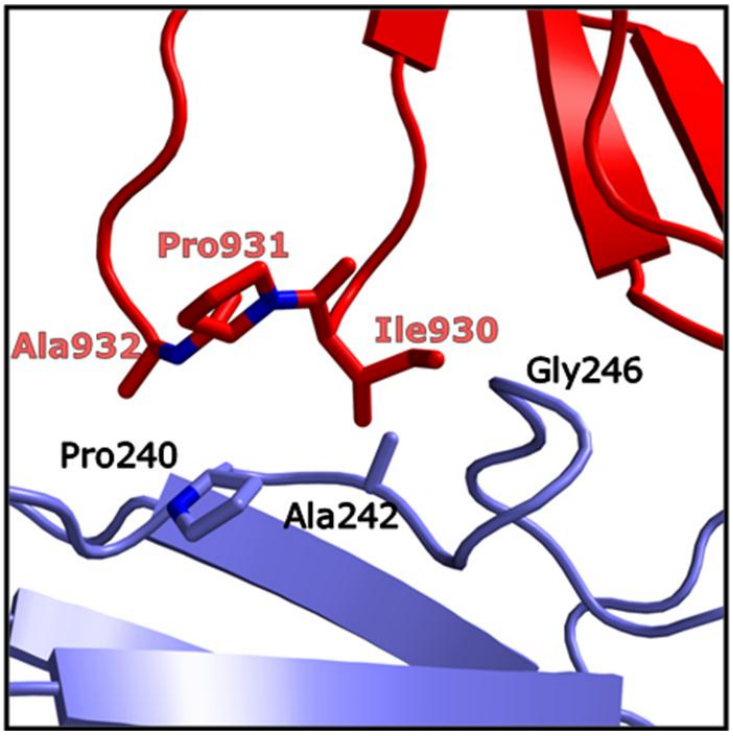


Figure 6.14: Detailed view onto binding site B. Residues involved in the Neo_{56}-RgmB interaction are represented as sticks. Colour coding as in Fig. 6.11.

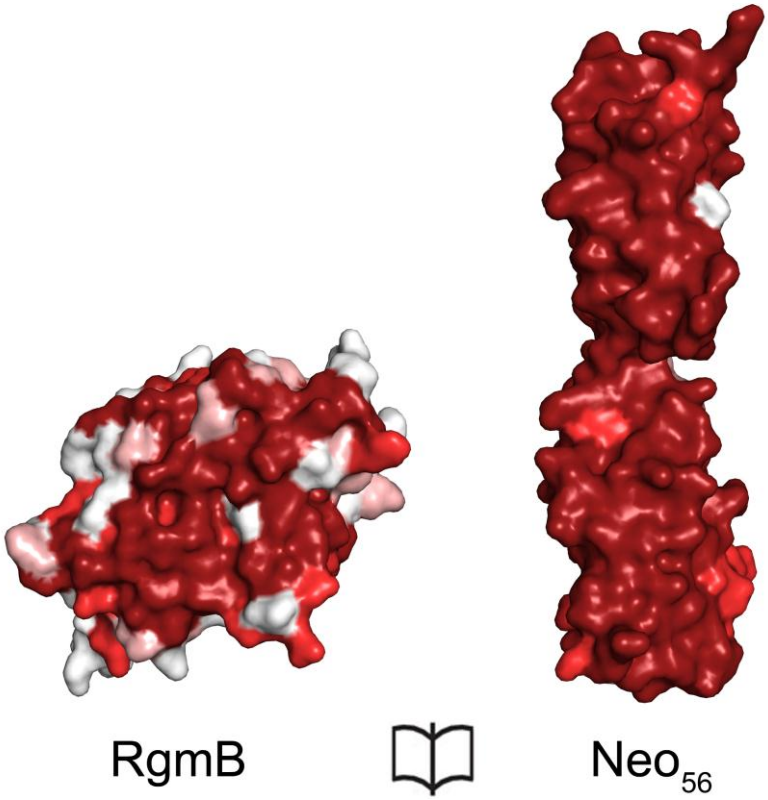


Figure 6.15: The Neo_{56}-RgmB interface is conserved. Surface representation of Neo_{56} and RgmB with sequence conservation coloured in white (0 %), salmon (>0 %), red (>33 %) and firebrick (>66 %).

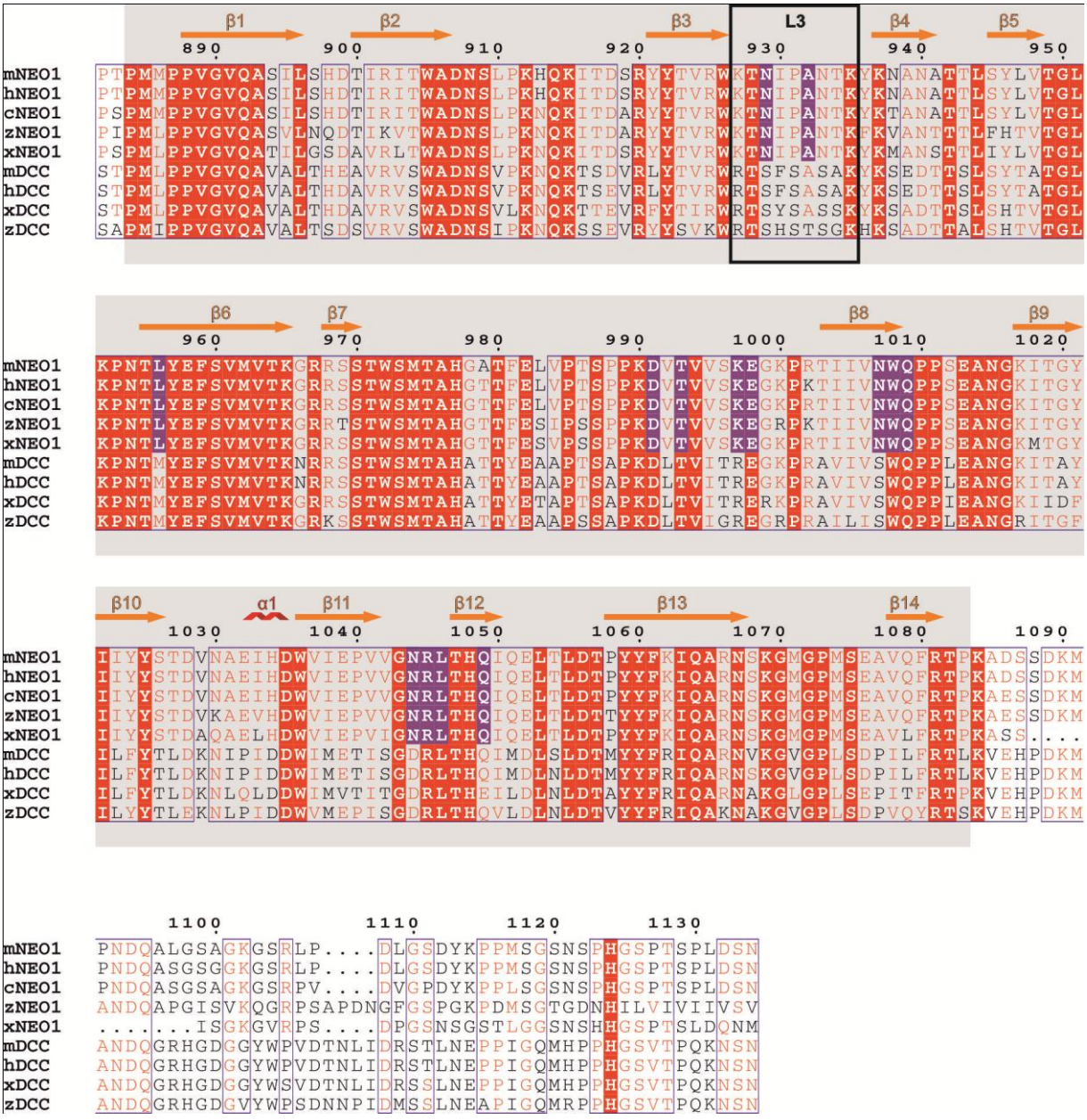


Figure 6.16: Sequence alignment of Neogenin and DCC from different species. The Neogenin sequences were aligned using MULTALIN (bioinfo.genotoul.fr/multalin/multalin.html) and formatted with ESPRIPT (espript.ibcp.fr/ESPript/ESPript/). Only the FN5 and FN6 domain as well as the membrane-proximal region are shown. Secondary structure elements are given above the mouse Neogenin sequence. The L3-loop region is marked with a box. Residues involved in the Neo56-RgmB interface are highlighted in purple.

To further validate the interface observed in the crystal structure, several interface mutants on both Neo_{56} and RgmB were designed, cloned and expressed. The binding of these mutants to RgmA, RgmB and RgmC (in case of Neo_{56} mutants) and Neo_{56} (in case of RgmB mutants) was then assessed by SPR (Table 7, Table 8, Fig. 6.17, Fig. 6.18, Fig. 6.19). The affinity between wild-type RgmB and Neo_{56} is 27 nM (Fig. 6.17B). Binding of RgmA and RgmC to Neo_{56} is 4-7 times stronger (Fig. 6.17A, Fig6.17C). A longer Neogenin construct, which included the O-link glycosylated, membrane proximal region (Neo_{56m}) showed a slightly weaker affinity for all Rgm family members, possibly due to the presence of the large sugars near the interface (Table 7).

Mutations in site A on either RgmB or Neo_{56} had a significant effect on their binding affinity (Fig. 6.17, Fig. 6.18, Fig. 6.19). Out of the 11 mutants that were studied, eight abolished binding completely whereas the remaining three showed an affinity that was at least two orders of magnitude weaker than the wild-type. In particular, introduction of an N-linked glycosylation site at residue 1007 in Neo_{56} lead to a complete loss of binding to all Rgm-family members, likely due to steric clashes of the sugar with Rgm. These results validate the interface that is observed in the crystal and underline the importance of site A for complex formation. Remarkably, most of the mutations showed similar effects across all three members of the Rgm-family, RgmA, RgmB and RgmC, respectively. This further strengthens the hypothesis that the Neo_{56}-RgmB interface described here reveals a common architecture for Rgm-Neogenin interactions.

Mutations in binding site B had a similarly dramatic effect on the affinity as site A mutants, however, none of the studied mutations led to a complete loss of binding. This is in full

agreement with the crystallographic analysis that suggested that site B is important but not essential for binding.

	K_d (nM)		
	RgmA	RgmB	RgmC
Neo_{56}	3.9 ± 0.4	27 ± 2	8 ± 1
Neo_{56m}	11 ± 1	48 ± 5	15 ± 2
^{930}GluGluGlu^{932}	80 ± 10	2880 ± 70	800 ± 200
Leu1046Glu	ND	ND	ND
Leu1046Arg	3000 ± 700	ND	ND
Gln1009Arg	1600 ± 200	ND	1300 ± 200
Gln1009Ser, Pro1010Ala	ND	ND	ND
Thr993Glu	ND	ND	ND
Val995Glu	190 ± 20	3300 ± 400	70 ± 10
Neo_{ecto}	39 ± 7	290 ± 60	70 ± 10
Neo_{ecto} Leu1046Glu	ND	ND	ND

Table 7: Binding affinities for different Neogenin constructs and mutants to Rgms. RgmA, RgmB and RgmC were immobilized on the Biacore chip, whereas the Neogenin constructs were used as analyte. All experiments were performed in duplicate. Binding constants (K_d) are given as mean with the error representing the standard error of the mean. All mutants were designed in the Neo_{56} background unless stated otherwise. ND: not determinable.

	K_d (nM)
Ala242Glu	3400 ± 100
Gln198Ala	3020 ± 60
Gly246Trp	1400 ± 200
Ala186Arg	ND
Pro240Glu	1100 ± 100
Asp191Arg	8000 ± 200
Gly246Glu	10400 ± 200
Thr217Glu	ND
Ile219Arg	ND

Table 8: Binding affinities for different RgmB mutants to Neo_{56}. The RgmB mutants were immobilized on the Biacore chip and Neo56 was used as analyte. Binding constants (K_d) are given as mean with the error representing the standard error of the mean. ND: not determinable.

In order to confirm that the interaction of Rgm with the Neogenin receptor is solely based on the interface observed in the crystal structure, the binding affinity of the full ectodomain of Neogenin (Neo_{ecto}) to RgmA, RgmB and RgmC, respectively, was studied (Fig. 6.18). Since the N- and C-termini of RgmB are disordered in the crystal, there was a possibility that these domains could interact with other parts of the Neogenin receptor that are not present in the Neo_{56} construct. The affinity of the Rgms to Neo_{ecto} is one order of magnitude weaker compared to the Neo_{56} construct. A possible reason for this could be the accessibility of the FN5 and FN6 domains in the context of the full receptor ectodomain. Introduction of the Leu1046Glu mutation that abolished binding of Rgms to Neo_{56} also led to a complete loss of

affinity for Neo_{ecto}. This result shows that the interface observed in the crystal structure presented here is exclusively responsible for the interaction between Rgm and Neogenin. The N- and C-termini of Rgm or the Neogenin domains not present in the Neo_{56} construct do not participate in the interaction. Therefore the structure presented here likely represents a complete description of the molecular determinants of the interaction of Rgm and its receptor Neogenin.

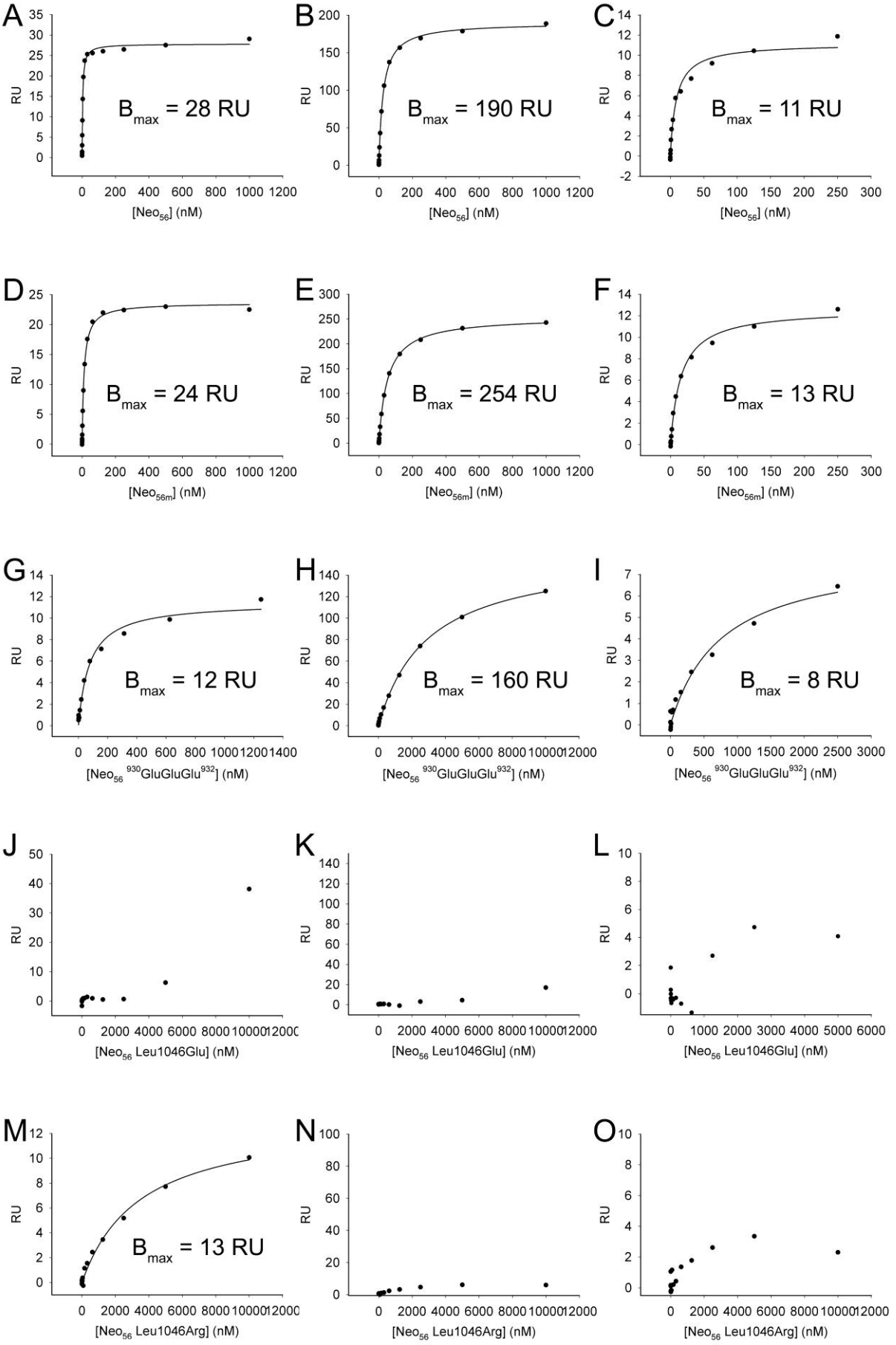
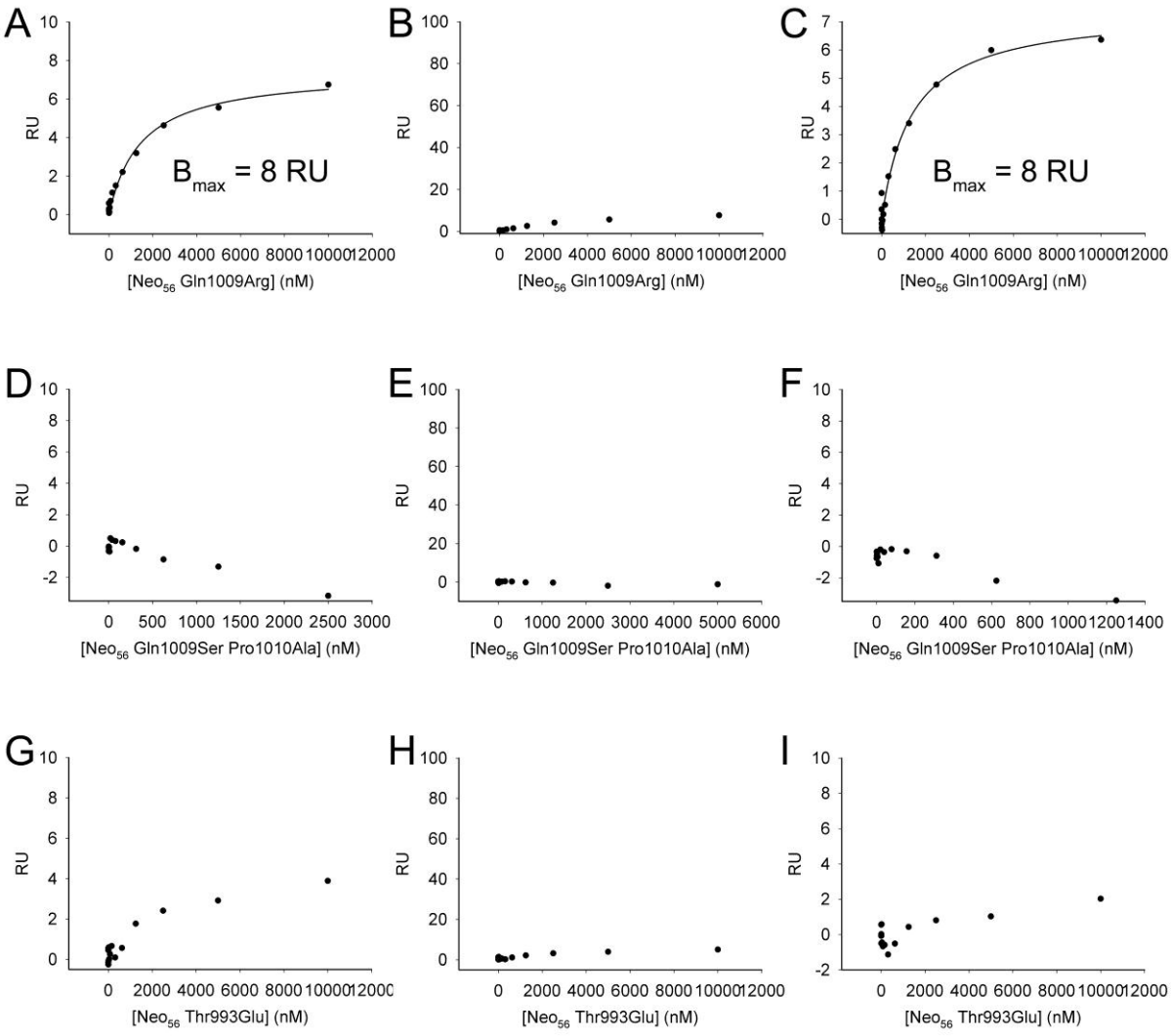

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
RU
B_{max} = 28 RU
B_{max} = 190 RU
B_{max} = 11 RU
B_{max} = 24 RU
B_{max} = 254 RU
B_{max} = 13 RU
B_{max} = 12 RU
B_{max} = 160 RU
B_{max} = 8 RU
B_{max} = 13 RU
[Neo_{56}] (nM)
[Neo_{56m}] (nM)
[Neo_{56} ^{930}GluGluGlu^{932}] (nM)
[Neo_{56} Leu1046Glu] (nM)
[Neo_{56} Leu1046Arg] (nM)

Figure 6.17: Binding of Neo_{56} site A mutants to RgmA, RgmB and RgmC, respectively. Graphs show a plot of the equilibrium binding response (response units (RU)) against Neo_{56} concentration ranging from 0.12 nM to 10 µM. Within one experiment each concentration was measured twice. Best-fit binding curves corresponding with a 1:1 binding model are shown as lines. The chip concentrations of the Rgm ligands were 80 RU, 500 RU and 90 RU for RgmA, RgmB and RgmC, respectively. (A) Neo_{56}-RgmA; (B) Neo_{56}-RgmB; (C) Neo_{56}-RgmC; (D) Neo_{56m}-RgmA; (E) Neo_{56m}-RgmB; (F) Neo_{56m}-RgmC; (G) Neo_{56} ^{930}GluGluGlu^{932}-RgmA; (H) Neo_{56} ^{930}GluGluGlu^{932}-RgmB; (I) Neo_{56} ^{930}GluGluGlu^{932}-RgmC; (J) Neo_{56} Leu1046Glu-RgmA; (K) Neo_{56} Leu1046Glu-RgmB; (L) Neo_{56} Leu1046Glu-RgmC; (M) Neo_{56} Leu1046Arg-RgmA; (N) Neo_{56} Leu1046Arg-RgmB; (O) Neo_{56} Leu1046Arg-RgmC. Corresponding K_d values can be found in Table 5.



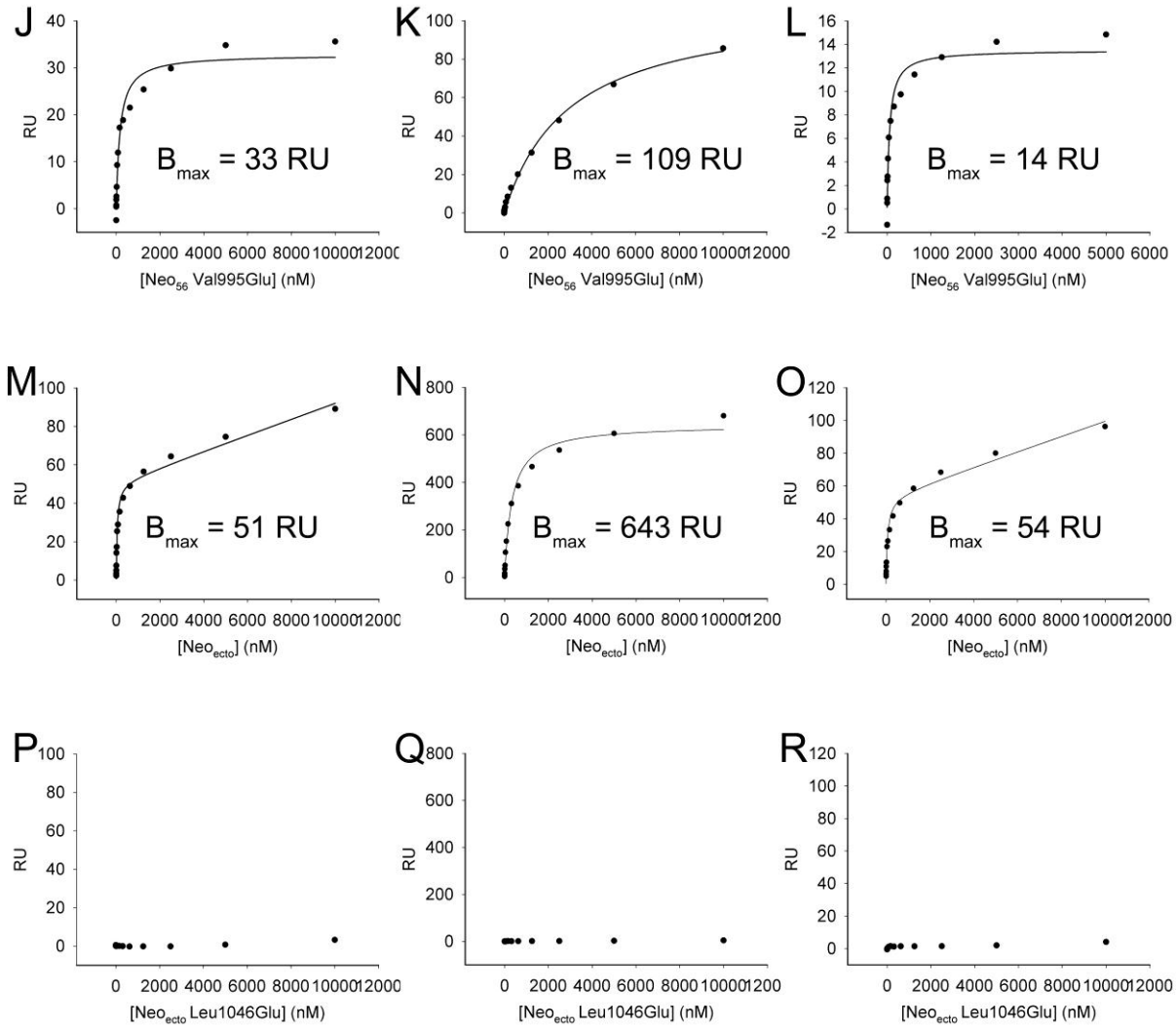


Figure 6.18: Binding of different Neogenin constructs and mutants to RgmA, RgmB and RgmC, respectively. Graphs show a plot of the equilibrium binding response (response units (RU)) against Neogenin concentration ranging from 0.12 nM to 10 µM. Within one experiment each concentration was measured twice. Best-fit binding curves corresponding with a 1:1 binding model are shown as lines. For Neo_{ecto}-RgmA (M) and Neo_{ecto}-RgmC (O) the data points were fitted with a 1:1 binding model that includes unspecific binding. The chip concentrations of the Rgm ligands were 80 RU, 500 RU and 90 RU for RgmA, RgmB and RgmC, respectively. (A) Neo_{56} Gln1009Arg-RgmA; (B) Neo_{56} Gln1009Arg-RgmB; (C) Neo_{56} Gln1009Arg-RgmC; (D) Neo_{56} Gln1009Ser,Pro1010Ala-RgmA; (E) Neo_{56} Gln1009Ser,Pro1010Ala-RgmB; (F) Neo_{56} Gln1009Ser,Pro1010Ala-RgmC; (G) Neo_{56} Thr993Glu-RgmA; (H) Neo_{56} Thr993Glu-RgmB; (I) Neo_{56} Thr993Glu-RgmC; (J) Neo_{56} Val995Glu-RgmA; (K) Neo_{56} Val995Glu-RgmB; (L) Neo_{56} Val995Glu-RgmC; (M) Neo_{ecto}-RgmA; (N) Neo_{ecto}-RgmB; (O) Neo_{ecto}-RgmC; (P) Neo_{ecto} Leu1046Glu-RgmA; (Q) Neo_{ecto} Leu1046Glu-RgmB; (R) Neo_{ecto} Leu1046Glu-RgmC. Corresponding K_d values can be found in Table 5.

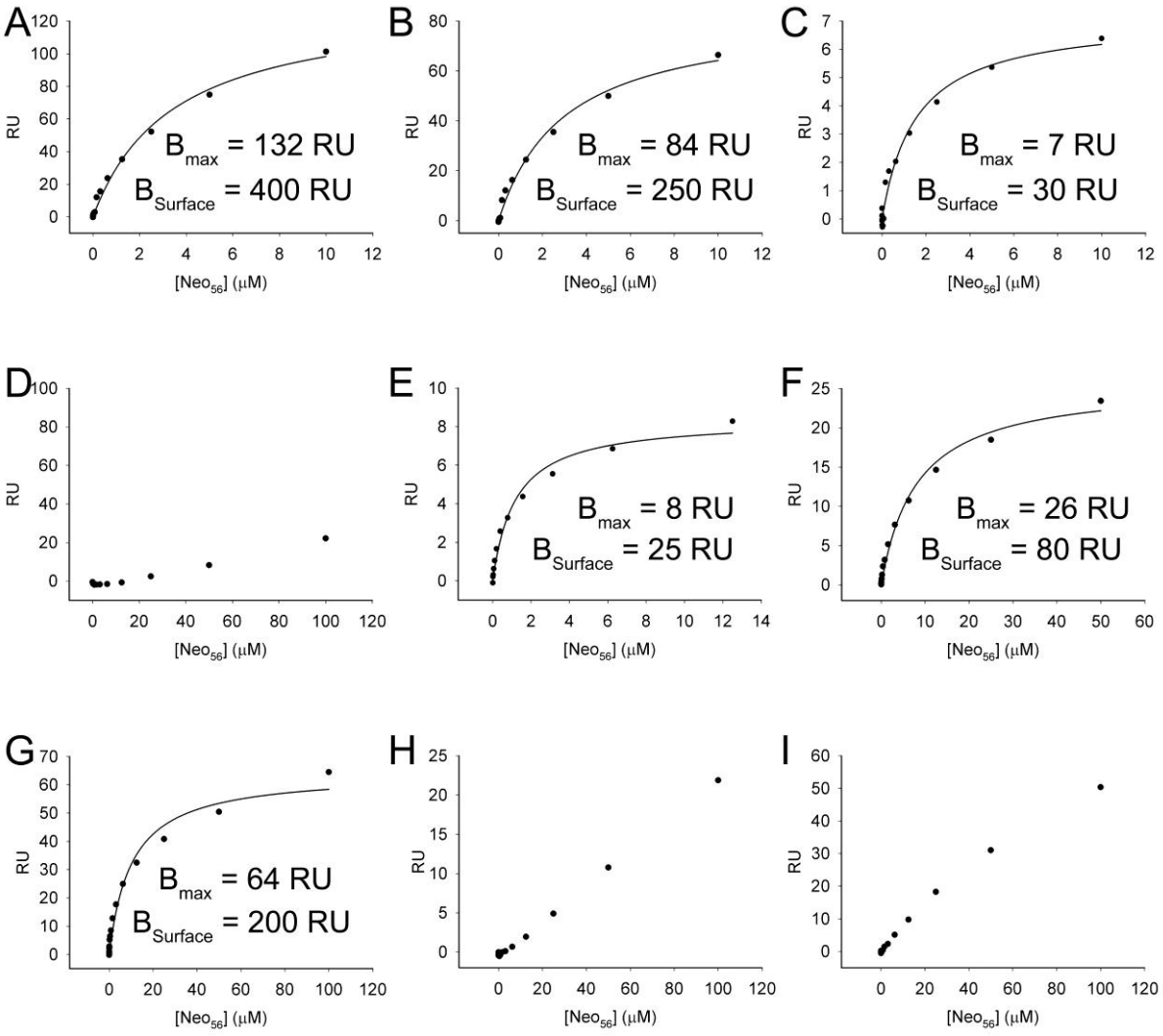


Figure 6.19: Binding of Neo_{56} to different RgmB mutants. Graphs show a plot of the equilibrium binding response (response units (RU)) against Neo_{56} concentration ranging from 6 nM to 100 µM. Within one experiment each concentration was measured twice. Best-fit binding curves corresponding with a 1:1 binding model are shown as lines. The chip concentrations of the RgmB ligands are given as $B_{surface}$.(A) RgmB Ala242Glu; (B) RgmB Gln198Ala; (C) RgmB Gly246Trp; (D) RgmB Ala186Arg; (E) RgmB Pro240Glu; (F) RgmB Asp191Arg; (G) RgmB Gly246Glu; (H) RgmB Thr217Glu; (I) RgmB Ile219Arg. Corresponding K_d values can be found in Table 6.

6.7 Structural and functional implications for the role of RgmB in BMP signalling

Members of the Rgm-family can act as co-receptors for BMP and promote activation of BMP-induced signalling cascades. Mutations in RgmC are directly linked to juvenile hemachromatosis, an iron disorder that is crucially dependent on BMP-induced hepcidin expression. In order to study the role of RgmB in BMP signalling and in particular the RgmB-BMP-2 interaction, it was necessary to produce functional BMP-2 in large quantities. BMP-2 was chosen since it is one of the best studied members of the BMP family and was shown to bind to Rgms [233]. Expression trials in mammalian cells did not yield any secreted, soluble protein, thus the protein was refolded from inclusion bodies (see Section 2.2.1). Once refolded, the biological activity of BMP-2 was assessed using a BMP-response assay with a Luciferase read-out (see Section 2.12). The refolded BMP-2 showed a dose-dependent response and at the highest concentration of 100 µM luminescence was increased by two orders of magnitude compared to the negative control (Fig. 6.20). Commercially available BMP-2 (R&D systems) usually generates a 15-fold increase in luminescence (personal communication Benjamin Bishop). In addition, the binding affinities of BMP-2 to its receptors AcvRII, BMPRII and BMPRIA were determined using SPR (Fig. 6.21). The binding constants lie in the low micromolar range for the type II receptors AcvRII and BMPRII and in the high nanomolar range for BMPRIA. This is in good agreement with previously published binding constants [234,235] and further validates the high quality of the refolded protein.

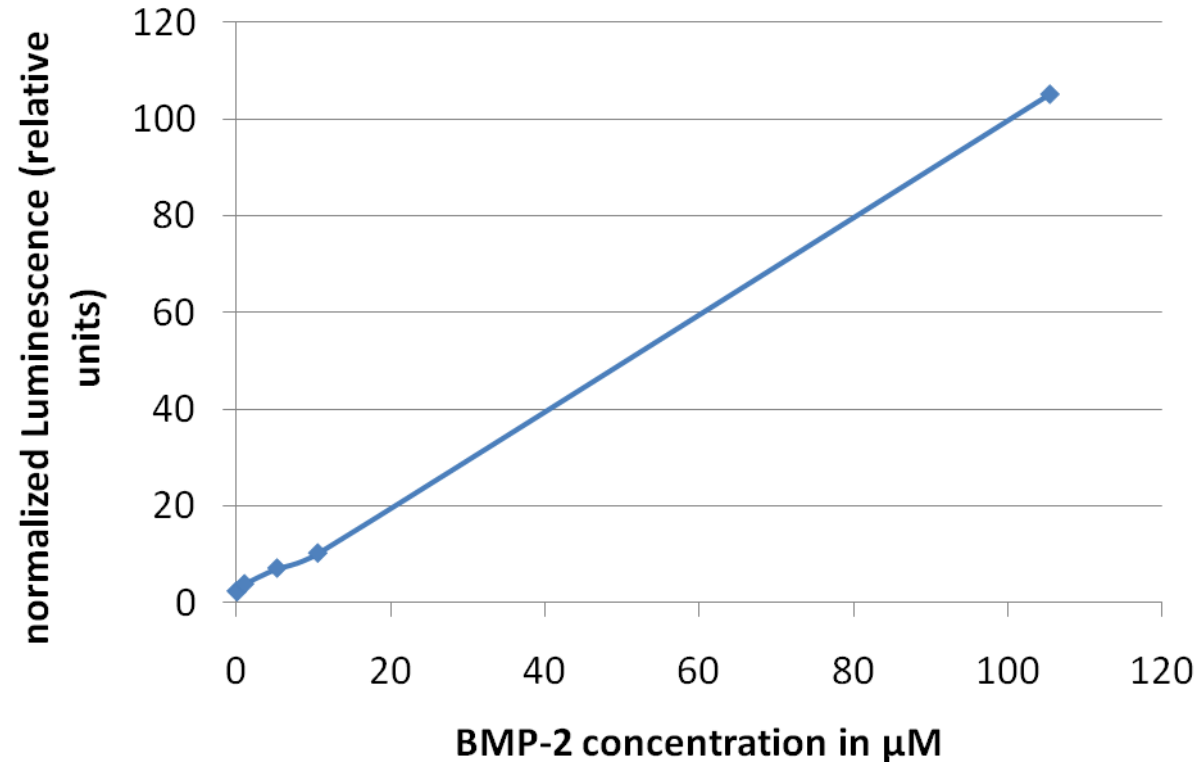


Figure 6.20: BMP response assay. Refolded BMP-2 was titrated to C2C12 cells as described in Section 2.12 and after cell lysis and addition of the luciferase substrate the relative luminescence was measured.

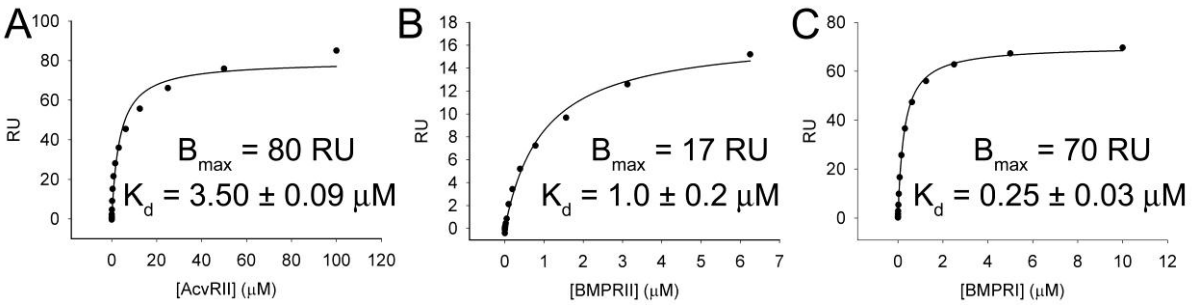


Figure 6.21: Binding of AcvRII, BMPRII and BMPRIA to BMP. Graphs show a plot of the equilibrium binding response (response units (RU)) against the receptor concentration ranging from 6 nM to 100 µM for AcvRII and BMPRII and from 0.6 nM to 10 µM for BMPRIA. Within one experiment each concentration was measured twice. Best-fit binding curves corresponding with a 1:1 binding model are shown as lines. The chip concentrations of the BMP ligand was 180RU. (A) AcvRII; (B) BMPRII; (C) BMPRIA.

In section 6.5 expression trials were carried out to test for expression of the so far reported disease related RgmC mutants. Only two of these loss-of-function mutants, Gly99Arg and Leu101Pro were secreted (Fig. 6.9). In order to test if their phenotype results from an altered affinity to BMP, an RgmB construct was designed that is missing the N-terminal domain (residues 137-411, RgmBΔN). This construct could be solubly expressed and purified in large scale. Wild-type RgmB binds to BMP-2 with an affinity of 47 nM whereas the binding affinity of the shorter RgmBΔN construct to BMP-2 is almost 200-fold weaker (Fig. 6.22). This suggests that at least the major BMP binding site on RgmB resides on the N-terminal domain and thus includes Gly99 and Leu101.

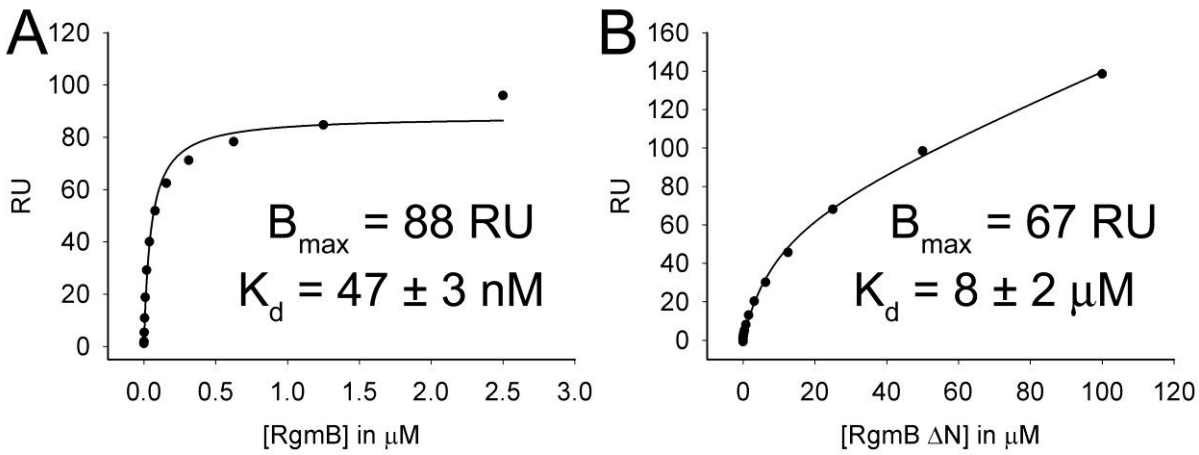


Figure 6.22: Binding of RgmB and RgmBΔN to BMP. Graphs show a plot of the equilibrium binding response (response units (RU)) against the RgmB concentration ranging from 0.6 nM to 10 µM for RgmB and from 6 nM to 100 µM for RgmBΔN. Within one experiment each concentration was measured twice. Best-fit binding curves corresponding with a 1:1 binding model are shown as lines. The chip concentrations of the BMP ligand was 180RU. (A) RgmB; (B) RgmBΔN.

6.8 Discussion

Members of the Rgm family are involved in cell migration, T-cell activation and iron homeostasis {Monnier, 2002 #87; Mirakaj, #93; Muramatsu, #124; }. The structural basis of Rgms that enables this diverse range of functions has so far remained unknown. Here the structure of RgmB in complex with its receptor Neogenin is presented and reveals a novel protein fold for the middle domain of RgmB. An autocatalytic Asp-Pro cleavage site is essential for the structural integrity of this domain. The middle domain of RgmB shows some sequence similarity with vWfd-domains, however it is missing the conserved GLCG motif. Thus, the middle domain of RgmB can be classified as vWfd-like domain. There is very little sequence identity among different vWfd-domain family members, therefore it is difficult to predict if the other family members will adopt a similar fold than the middle domain of RgmB. Interestingly, some but not all family members share the Asp-Pro cleavage site and Mucin and BMPER have been reported to undergo autocatalytic cleavage [225,236]. This suggests that the acid-labile, cleavable Asp-Pro bond might be a common posttranslational modification that increases the structural integrity of vWfd-domains.

The interface between RgmB and Neogenin in the complex structure has been further characterized by extensive mutagenesis and biophysical analysis using SPR. The interaction is exclusively mediated by the middle domain of RgmB and the FN5 and FN6 domain of Neogenin. DCC is a Netrin-receptor that shows high sequence similarity to Neogenin in particular in the fourth, fifth and sixth FNIII domain [104]. However, previous studies revealed that DCC does not bind Rgms [104]. The FN6 domain of DCC and Neogenin share high sequence identity in particular in the interface region. In contrast, the L3 loop of the FN5

domain that is crucially involved in the RgmB-Neo56 interaction shows almost no identity. This suggests that the L3 loop might play an important role in receptor recognition and indeed function. To further validate this hypothesis, several Neo_{56} apo structures were superposed onto the Neo_{56} structure in the complex with RgmB (Fig. 6.22). There is a large degree of flexibility between the FN5 and FN6 domain. Interaction with RgmB locks the FN5-FN6 domain hinge in a particular orientation by interaction of RgmB with the L3 loop region. This loop region can only be observed in the complex structure suggesting that it is ordered upon complex formation. The FN5-FN6 hinge stabilization could be essential for triggering downstream Neogenin signalling events and for mediating Rgm-induced Neogenin activation. Future work will need to focus on this possibility and in particular the importance of the L3 loop. Data presented here strongly suggests a regulatory and possibly even mechanistic role for this loop region.

The biophysical characterization not only validated the interface observed in the crystal structure but also revealed that mutations in the interface have similar effects on the binding of Neogenin to RgmA, RgmB and RgmC, respectively. Indeed, the interface residues on both proteins are conserved among different species and Rgm-family members suggesting a common binding mode for all Rgm-Neogenin interactions. Since the different Rgm constructs were immobilized on the Biacore chip, their activity can be determined since ligand activity (ActL), stoichiometry (S), molecular mass of ligand (M_L) and anaylte (M_A), the level of immobilised ligand (L) and the analyte binding level at saturation (A) are related as follows:

$$ActL*S = M_L*M_A^{-1}*A*L^{-1}$$

RgmA showed an activity of ~45 %, RgmB ~70 % and RgmC ~25 %. The apparent lack in activity can be explained by the protein preparation prior to loading on the chip. The proteins are buffer-exchanged into a buffer containing no salt and no purification step is carried out. Thus, protein aggregates that are not able to bind to Neogenin but still carry an accessible biotinylation sequence will be deposited on the chip but will not contribute to binding. All RgmB mutants but Gly246Trp showed a similar activity to the wild-type protein. The Gly246Trp mutation likely exposes the hydrophobic Trp and facilitates aggregation.

Mutations on RgmC have been reported to be directly linked to juvenile hemachromatosis, an iron disorder in which BMP-regulated hepcidin expression is downregulated [215,216]. Rgms act as co-receptors for BMP and increase BMP signalling. In this chapter all of the identified RgmC mutations were studied in order to find a rationale for their phenotype. Only two mutants, Gly99Arg and Leu101Pro, were solubly expressed. All other mutants were not secreted in the mammalian expression system and thus cannot be functional. The two secreted mutations map onto the N-terminal domain of RgmB that could not be observed in the complex structure. SPR analysis showed that this domain is responsible for BMP binding suggesting that Gly99Arg and Leu101Pro are either directly involved in the Rgm-BMP interface or alter the local folding of Rgm so that it cannot interact with BMP any longer. Since the Neogenin binding site on RgmB is distant from the N-terminal domain and does not interact with it as judged from the crystal structure, concomitant binding of Neogenin and BMP to Rgm seems possible as proposed by Zhang et al. and Zhou et al. in functional studies [111,217].

Finally, expression of RgmB did not suggest any homophilic interaction that could mediate cell-cell adhesion since both the protein itself and the complex with its receptor Neo_{56} behave as monomeric species on gelfiltration, even at high concentrations. However, additional work, in particular on a cellular level, is necessary to exclude this possibility for Rgm-function.

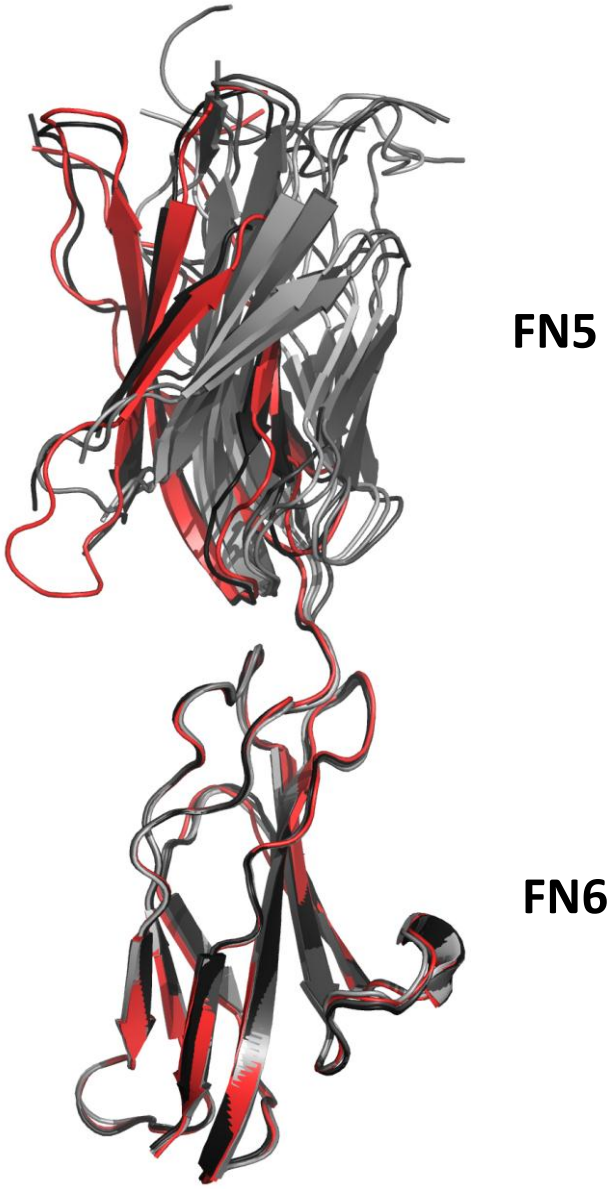


Figure 6.23: Superposition of Neo56 from different apo structures and the complex with RgmB. Coordinates for the apo structures were taken from Yang et al. (PDB ID 3P4L [74]) or were provided by Christian Siebold (unpublished). All structures were superimposed on the FN6 domain using SSM as implemented in COOT [124]. The apo structures are coloured in different shades of gray, the complex structure is shown in red.

Chapter 7

Conclusion

Chapter 7 Conclusion

Chemotaxis is a recurring theme in all kingdoms of life and despite different complexities of the underlying signal transduction pathways its basic mechanism is conserved. An extracellular stimulus is sensed by a transmembrane receptor which then alters the activation state of a molecular switch. This modification of the activity of the switch then causes further downstream processes, ultimately leading to a bias in cell motility.

The best understood chemotactic system is found in bacteria. Despite extensive research in the field it was unclear how some bacteria manage to have over 250 different and independent two-component signalling systems. In this thesis structural characterization of the interaction between a sensor kinase and its cognate response regulator allowed the specificity generating regions to be pinpointed which then allowed generation of novel two-component pairs. Understanding the intricate regulation of distinct signalling partners is essential to be able to design synthetic two-component signalling circuits.

Chemotactic signal transduction is dependent on the ability of molecular switches, like response regulators and small GTPases, to change between active and inactive conformation very quickly. The molecular mechanism of this change has been studied extensively in both prokaryotes and eukaryotes. The structural and functional characterization of a response regulator from *R. sphaeroides* described in this thesis provides a novel view onto the activation mechanism and its molecular determinants. However, these insights need to be validated further by ongoing work on the dynamics of response regulator activation.

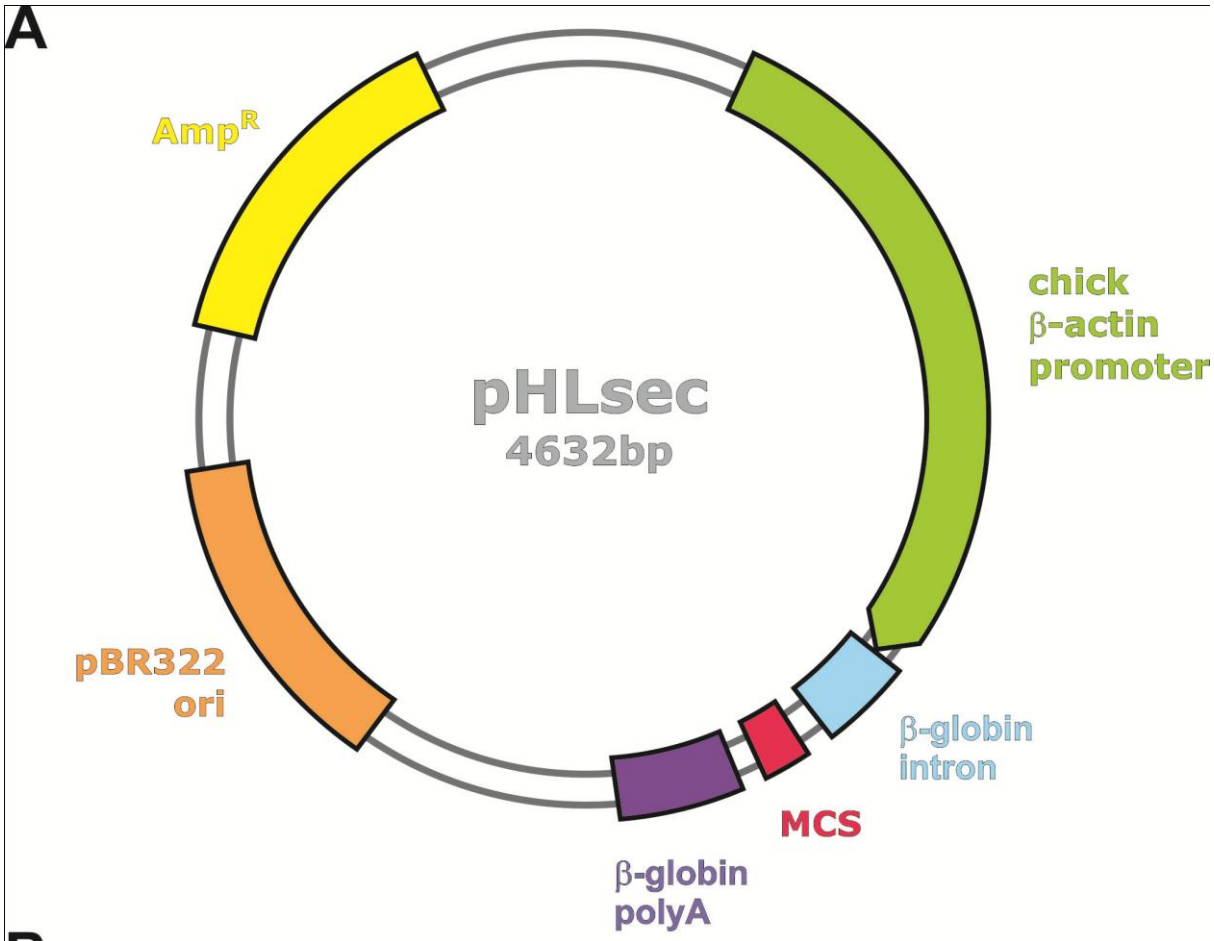
In eukaryotic molecular switches, small GTPases, the intrinsic activation and inactivation is too slow to cope with the need for a fast response to an incoming signal. Therefore, regulatory proteins are needed to catalyze both reactions. The intracellular region of plexins inactivates small GTPases of the Ras family, in the presence of both the extracellular as well as intracellular ligand. The molecular mechanism of this complicated regulation has remained unclear. Structural and functional data presented in this thesis now suggest a model that elegantly combines the binding events on the outside and inside of the cell into a single signalling output.

Like plexins, Rgms play an important role in axon guidance. No structural information on these proteins was available that could facilitate explanation of their functional role and their receptor interaction. Here, the structure of the major domain of Rgm in complex with its receptor is presented and together with biophysical and functional data provides a rationale for its role in the disease juvenile hemachromatosis. Further work is needed to deduct the exact mechanism by which Rgms can activate their receptor Neogenin.

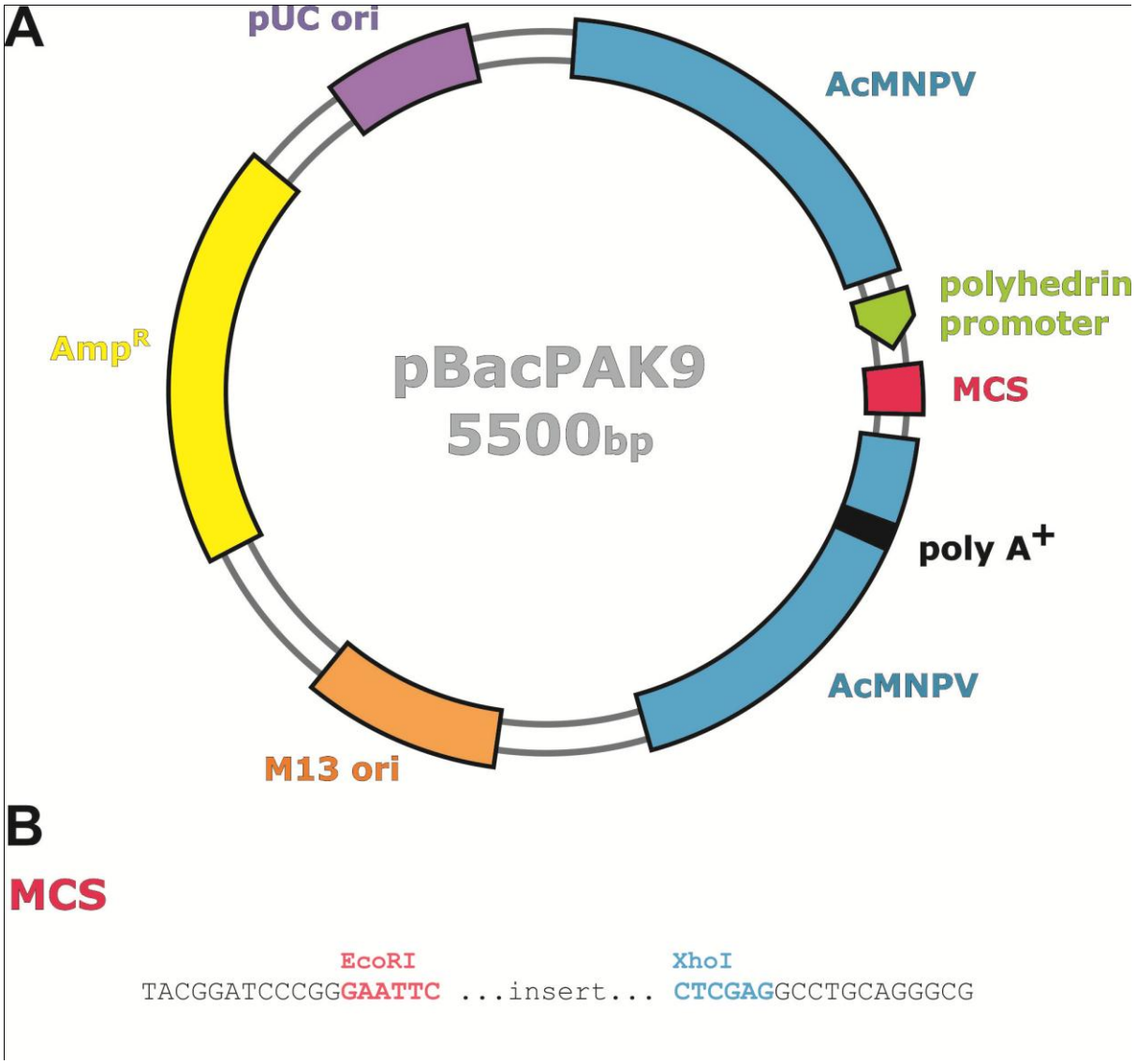
In summary, several components of chemotactic pathways in both prokaryotes and eukaryotes were studied in this thesis. In all cases structural information from X-ray crystallography was combined with biophysical and functional assays suggesting specific signalling or activation mechanisms that had not been described previously. Proteins were studied on their own and in complex with their binding partners and included intracellular as well as secreted molecules produced in bacteria, insect cells and mammalian cells. Despite the overall disparity of these works, they unify under the theme of using structural biology to generate novel insights into chemotactic signal transduction.

Appendices

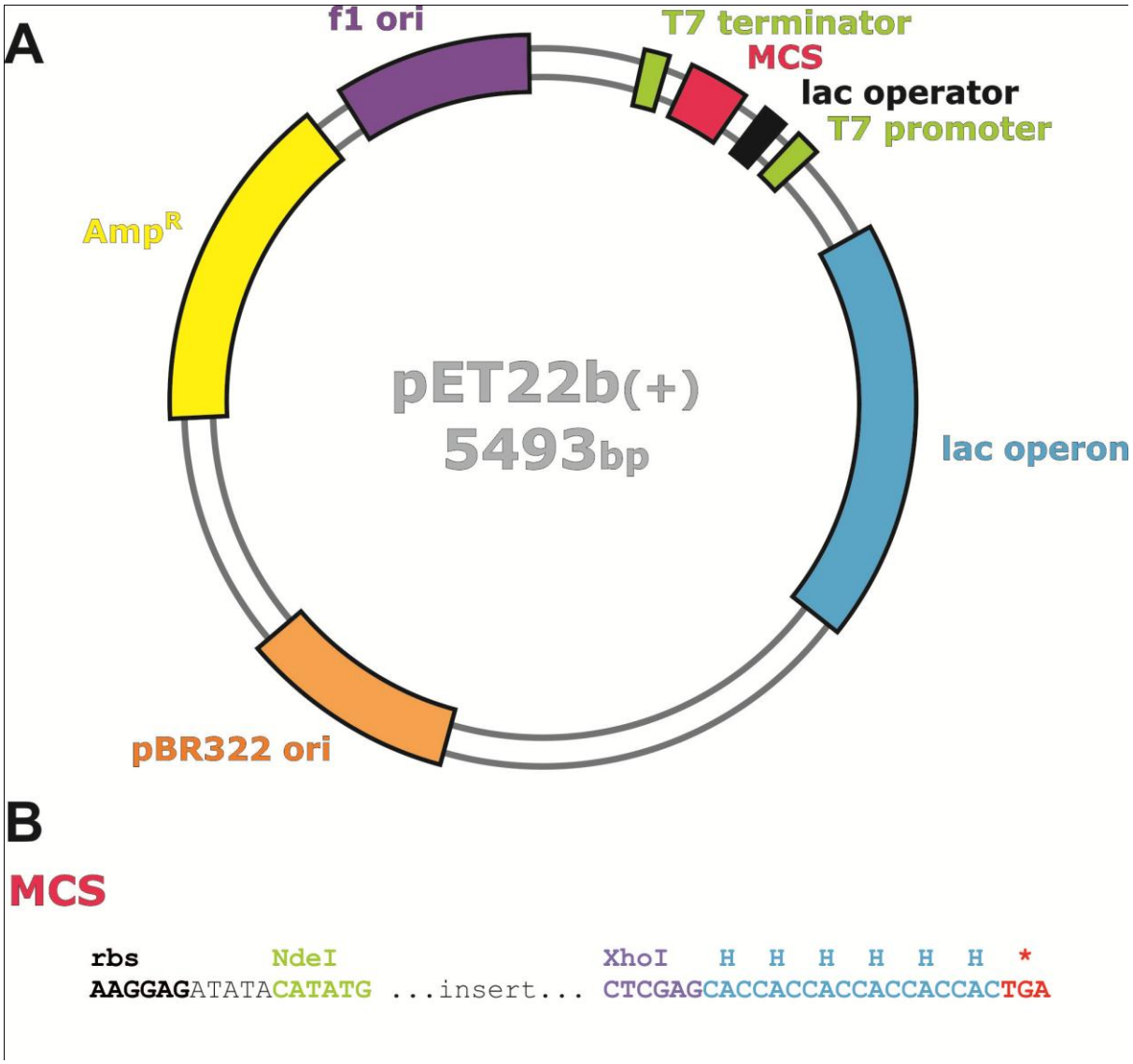
Appendix A Expression vectors



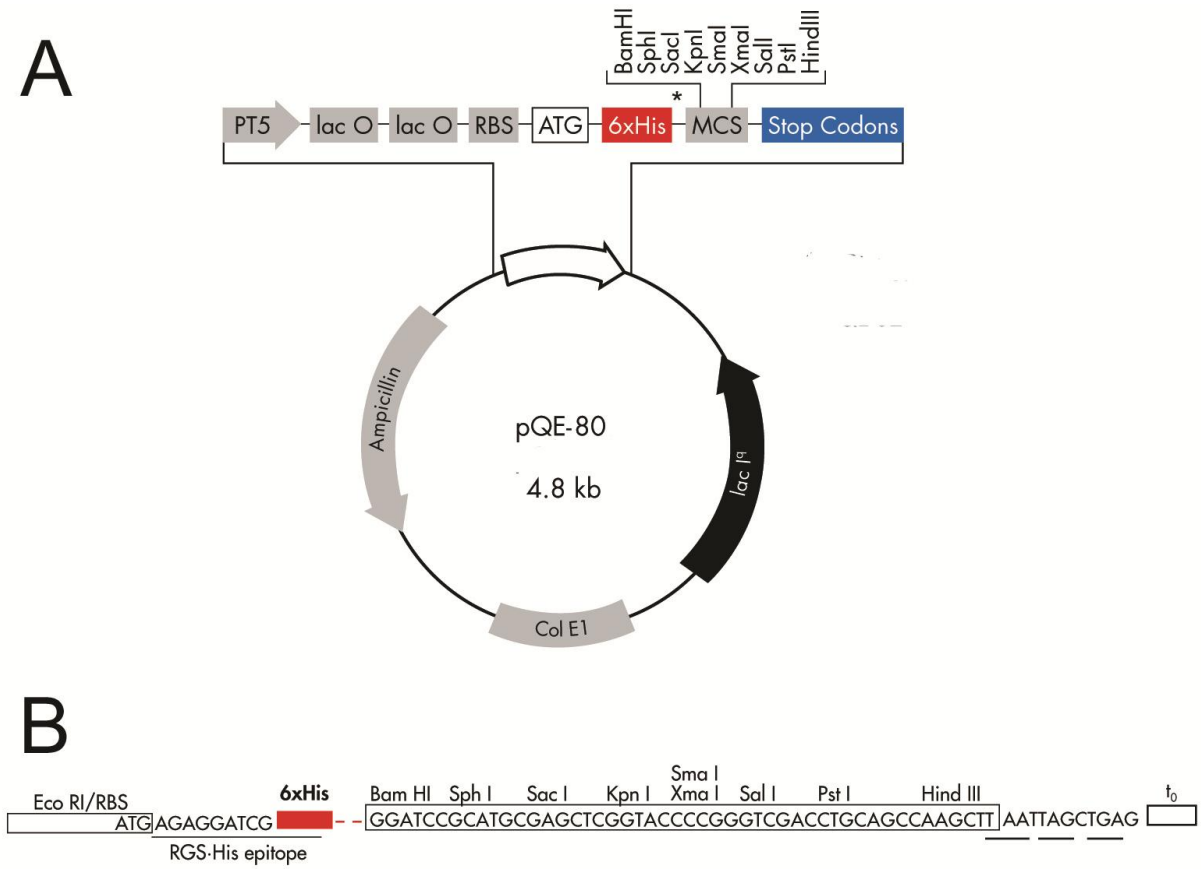
Key features of the pHLSec vector. (A) Vector map of pHLSec. (B) Multiple cloning site of pHLSec. The N-terminal Kozak sequence is highlighted in black and the chicken RPTPσ secretion signal in green, the C-terminal His_6-tag is shown in blue and stop codons in red, while specific restriction enzyme sites are colour coded. Figure based on Aricescu et al. [112].



Key features of the pBacPAK9 vector (Clontech). (A) Vector map of pBacPAK9. (B) Restriction enzyme sites used for cloning into the multiple cloning site of pBacPAK9 are highlighted.



Key features of the pET22b (+) vector (Invitrogen). (A) Vector map of pET22b (+). (B) Selected restriction enzyme sites are colour coded. The ribosome binding site is shown in black, the C-terminal His_6 tag in blue and the stop codon in red.



Key features of the pQE-80 vector (Qiagen). (A) Vector map of pQE80. (B) Multiple cloning site. Figure adapted from http://www.qiagen.com/literature/pqesequences/pqe8x.pdf.

Appendix B Abbreviations

µg Microgram

µL Microlitre

µM Micromolar

µm Micrometre

asu Asymmetric Unit

ATP Adenosine Triphosphate

BirA Biotin-[acetyl-CoA carboxylase] Synthetase

C Celsius

CCD Charge-coupled Device

CCP4 Collaborative Computing Project 4

cDNA Complementary Deoxyribonucleic Acid

CM5 Carboxymethyl 5

CNS Crystallography and NMR System

Da Dalton

DMEM Dulbecco's Modified Eagles Medium

DMSO Dimethylsulfoxide

DNA Deoxyribonucleic Acid

dNTP Deoxyribonucleotide Triphosphate

DSSP Dictionary of Protein Secondary Structure

ECL Enhanced Chemiluminescence

EDC *N*-ethyl-*N*‟-(3-dimethylaminopropyl)-carbodimide Hydrochloride

EDTA Ethylenediamine Tetra-acetic Acid

EndoF1 Endo-*N*-acetylglucosaminidase F1

ESRF European Synchrotron Radiation Facility

FBS Fetal Bovine Serum

FOM Figure of Merit

FPLC Fast Protein Liquid Chromatography

GlcNAc *N*-acetyleglucosamine

GPI Glycosylphosphatidylinositol

HEK-293T Human Embryonic Kidney 293 cell line expressing SV40 type T antigen

HEPES 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic Acid Hemisodium Salt

His_6-tag Hexa Histidine Tag

HK Histidine kinase

Hpt Histidine phosphotransfer domain

IMAC Immobilised Metal Affinity Chromatography

K Kelvin

kb Kilobase

K_d Dissociation constant

kDa Kilodalton

Kif Kifunensine

L Litre

LB Luria-Bertani Media

LBD Ligand Binding Domain

LLG Log-likelihood Gain Appendix A. Abbreviations 199

M Molar

MAN Mannose

MES 2-(*N*-Morpholino)ethanesulfonic Acid

mg Milligram

mM Millimolar

MOPS 3-(*N*-morpholino)propanesulfonic Acid

mRNA Messenger Ribonucleic Acid

NCS Non-crystallographic Symmetry

NHS *N*-hydroxysuccinimide

nL Nanolitre

nm Nanometre

nM Nanomolar

OD Optical Density of Monochromatic Light

P Phosphoprotein

PBS Phosphate Buffer Saline

PCR Polymerase Chain Reaction

PDB Protein Data Bank

PDZ PSD-95-DlG/ZO-1

PEG Polyethylene Glycol

PEI Polyethylenimine

pg Picogram

PISA Protein Interfaces Surfaces and Assemblies

RBD RhoGTPase Binding Domain

RMS Root Mean Square

RNA Ribonucleic Acid

RONN Regional Order Neural Network

RPTPσ Receptor Protein Tyrosine Phosphatase σ

RR Response regulator

RU Response Unit

s Seconds

SDS-PAGE Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

SEC Size Exclusion Chromatography

SHP Structure Homology Program

SPR Surface Plasmon Resonance

TAE Tris Acetate EDTA

TLS Translation-libration-screw

TLSMD TLS Motion Determination

TRIS 2-Amino-2-hydroxymethyl-1,3-propanediol

U Unit

V Volt

v/v Volume by volume

w/v weight by volume

Appendix C Publications

Publications arising from work performed in this thesis are listed below.

1. **Bell, C.H.**, Porter, S.L. Strawson, A., Stuart, D.I., Armitage, J.P. Using structural information to change the phosphotransfer specificity of a two-component chemotaxis signalling complex. *PLoS Biology* **8**:e1000306 (2010).

2. Janssen, B.J., Robinson, R.A., Perez-Branguli, F., **Bell, C.H.**, Mitchell, K.J., Siebold, C., Jones, E.Y. Structural basis of semaphorin-plexin signalling. *Nature*, **467**:1118-22 (2010).

3. **Bell, C.H.**, Aricescu, A.R., Jones, E.Y., Siebold, C. A dual binding mode for RhoGTPases in plexin signalling. *PLoS Biology* **9(8):**e1001134 (2011).

References

1. Williams, S.M. et al. Helicobacter pylori chemotaxis modulates inflammation and bacterium-gastric epithelium interactions in infected mice. *Infect Immun* **75**, 3747-57 (2007).
2. Garvis, S. et al. Caenorhabditis elegans semi-automated liquid screen reveals a specialized role for the chemotaxis gene cheB2 in Pseudomonas aeruginosa virulence. *PLoS Pathog* **5**, e1000540 (2009).
3. Kedrin, D., van Rheenen, J., Hernandez, L., Condeelis, J. & Segall, J.E. Cell motility and cytoskeletal regulation in invasion and metastasis. *J Mammary Gland Biol Neoplasia* **12**, 143-52 (2007).
4. Luster, A.D. Chemokines--chemotactic cytokines that mediate inflammation. *N Engl J Med* **338**, 436-45 (1998).
5. Berg, H.C. Bacterial flagellar motor. *Curr Biol* **18**, R689-91 (2008).
6. Hazelbauer, G.L., Falke, J.J. & Parkinson, J.S. Bacterial chemoreceptors: high-performance signaling in networked arrays. *Trends Biochem Sci* **33**, 9-19 (2008).
7. Maddock, J.R. & Shapiro, L. Polar location of the chemoreceptor complex in the Escherichia coli cell. *Science* **259**, 1717-23 (1993).
8. Bray, D., Levin, M.D. & Morton-Firth, C.J. Receptor clustering as a cellular mechanism to control sensitivity. *Nature* **393**, 85-8 (1998).
9. Sourjik, V. & Berg, H.C. Receptor sensitivity in bacterial chemotaxis. *Proc Natl Acad Sci U S A* **99**, 123-7 (2002).
10. Dyer, C.M., Vartanian, A.S., Zhou, H. & Dahlquist, F.W. A Molecular Mechanism of Bacterial Flagellar Motor Switching. *J Mol Biol* (2009).
11. Vladimirov, N. & Sourjik, V. Chemotaxis: how bacteria use memory. *Biol Chem* **390**, 1097-104 (2009).
12. Anand, G.S. & Stock, A.M. Kinetic basis for the stimulatory effect of phosphorylation on the methylesterase activity of CheB. *Biochemistry* **41**, 6752-60 (2002).
13. Rao, C.V., Glekas, G.D. & Ordal, G.W. The three adaptation systems of Bacillus subtilis chemotaxis. *Trends Microbiol* **16**, 480-7 (2008).
14. Silversmith, R.E. Auxiliary phosphatases in two-component signal transduction. *Curr Opin Microbiol* **13**, 177-83.
15. Gauden, D.E. & Armitage, J.P. Electron transport-dependent taxis in Rhodobacter sphaeroides. *J Bacteriol* **177**, 5853-9 (1995).
16. Mackenzie, C. et al. The home stretch, a first analysis of the nearly completed genome of Rhodobacter sphaeroides 2.4.1. *Photosynth Res* **70**, 19-41 (2001).
17. Mackenzie, C. et al. Postgenomic adventures with Rhodobacter sphaeroides. *Annu Rev Microbiol* **61**, 283-307 (2007).
18. Pilizota, T. et al. A molecular brake, not a clutch, stops the Rhodobacter sphaeroides flagellar motor. *Proc Natl Acad Sci U S A* **106**, 11582-7 (2009).
19. Nicolau, D.V., Jr., Armitage, J.P. & Maini, P.K. Directional persistence and the optimality of run-and-tumble chemotaxis. *Comput Biol Chem* **33**, 269-74 (2009).

20. Hamblin, P.A., Maguire, B.A., Grishanin, R.N. & Armitage, J.P. Evidence for two chemosensory pathways in Rhodobacter sphaeroides. *Mol Microbiol* **26**, 1083-96 (1997).
21. Jeziore-Sassoon, Y., Hamblin, P.A., Bootle-Wilbraham, C.A., Poole, P.S. & Armitage, J.P. Metabolism is required for chemotaxis to sugars in Rhodobacter sphaeroides. *Microbiology* **144 (Pt 1)**, 229-39 (1998).
22. Wadhams, G.H., Martin, A.C., Warren, A.V. & Armitage, J.P. Requirements for chemotaxis protein localization in Rhodobacter sphaeroides. *Mol Microbiol* **58**, 895-902 (2005).
23. Porter, S.L., Roberts, M.A., Manning, C.S. & Armitage, J.P. A bifunctional kinase-phosphatase in bacterial chemotaxis. *Proc Natl Acad Sci U S A* **105**, 18531-6 (2008).
24. Porter, S.L. & Armitage, J.P. Chemotaxis in Rhodobacter sphaeroides requires an atypical histidine protein kinase. *J Biol Chem* **279**, 54573-80 (2004).
25. Porter, S.L. et al. The CheYs of Rhodobacter sphaeroides. *J Biol Chem* **281**, 32694-704 (2006).
26. Ferre, A., De La Mora, J., Ballado, T., Camarena, L. & Dreyfus, G. Biochemical study of multiple CheY response regulators of the chemotactic pathway of Rhodobacter sphaeroides. *J Bacteriol* **186**, 5172-7 (2004).
27. Tindall, M.J., Porter, S.L., Maini, P.K. & Armitage, J.P. Modeling chemotaxis reveals the role of reversed phosphotransfer and a bi-functional kinase-phosphatase. *PLoS Comput Biol* **6**.
28. Swaney, K.F., Huang, C.H. & Devreotes, P.N. Eukaryotic chemotaxis: a network of signaling pathways controls motility, directional sensing, and polarity. *Annu Rev Biophys* **39**, 265-89.
29. Chung, C.Y., Lee, S., Briscoe, C., Ellsworth, C. & Firtel, R.A. Role of Rac in controlling the actin cytoskeleton and chemotaxis in motile cells. *Proc Natl Acad Sci U S A* **97**, 5225-30 (2000).
30. Killeen, M.T. & Sybingco, S.S. Netrin, Slit and Wnt receptors allow axons to choose the axis of migration. *Dev Biol* **323**, 143-51 (2008).
31. Kruger, R.P., Aurandt, J. & Guan, K.L. Semaphorins command cells to move. *Nat Rev Mol Cell Biol* **6**, 789-800 (2005).
32. Pitulescu, M.E. & Adams, R.H. Eph/ephrin molecules--a hub for signaling and endocytosis. *Genes Dev* **24**, 2480-92.
33. Quinn, C.C. & Wadsworth, W.G. Axon guidance: asymmetric signaling orients polarized outgrowth. *Trends Cell Biol* **18**, 597-603 (2008).
34. Matsunaga, E. & Chedotal, A. Repulsive guidance molecule/neogenin: a novel ligand-receptor system playing multiple roles in neural development. *Dev Growth Differ* **46**, 481-6 (2004).
35. Janssen, B.J. et al. Structural basis of semaphorin-plexin signalling. *Nature* **467**, 1118-22.
36. Liu, H. et al. Structural basis of semaphorin-plexin recognition and viral mimicry from Sema7A and A39R complexes with PlexinC1. *Cell* **142**, 749-61.
37. Nogi, T. et al. Structural basis for semaphorin signalling through the plexin receptor. *Nature* **467**, 1123-7.

38. Hohenester, E. Structural insight into Slit-Robo signalling. *Biochem Soc Trans* **36**, 251-6 (2008).
39. Seiradake, E., Harlos, K., Sutton, G., Aricescu, A.R. & Jones, E.Y. An extracellular steric seeding mechanism for Eph-ephrin signaling platform assembly. *Nat Struct Mol Biol* **17**, 398-402.
40. Negishi, M., Oinuma, I. & Katoh, H. Plexins: axon guidance and signal transduction. *Cell Mol Life Sci* **62**, 1363-71 (2005).
41. Neufeld, G. & Kessler, O. The semaphorins: versatile regulators of tumour progression and tumour angiogenesis. *Nat Rev Cancer* **8**, 632-45 (2008).
42. Gherardi, E., Love, C.A., Esnouf, R.M. & Jones, E.Y. The sema domain. *Curr Opin Struct Biol* **14**, 669-78 (2004).
43. He, H., Yang, T., Terman, J.R. & Zhang, X. Crystal structure of the plexin A3 intracellular region reveals an autoinhibited conformation through active site sequestration. *Proc Natl Acad Sci U S A* **106**, 15610-5 (2009).
44. Tong, Y. et al. Structure and function of the intracellular region of the plexin-b1 transmembrane receptor. *J Biol Chem* **284**, 35962-72 (2009).
45. He, Z. & Tessier-Lavigne, M. Neuropilin is a receptor for the axonal chemorepellent Semaphorin III. *Cell* **90**, 739-51 (1997).
46. Kolodkin, A.L. et al. Neuropilin is a semaphorin III receptor. *Cell* **90**, 753-62 (1997).
47. Barberis, D. et al. Plexin signaling hampers integrin-based adhesion, leading to Rho-kinase independent cell rounding, and inhibiting lamellipodia extension and cell motility. *FASEB J* **18**, 592-4 (2004).
48. Rohm, B., Rahim, B., Kleiber, B., Hovatta, I. & Puschel, A.W. The semaphorin 3A receptor may directly regulate the activity of small GTPases. *FEBS Lett* **486**, 68-72 (2000).
49. Oinuma, I., Ishikawa, Y., Katoh, H. & Negishi, M. The Semaphorin 4D receptor Plexin-B1 is a GTPase activating protein for R-Ras. *Science* **305**, 862-5 (2004).
50. Saito, Y., Oinuma, I., Fujimoto, S. & Negishi, M. Plexin-B1 is a GTPase activating protein for M-Ras, remodelling dendrite morphology. *EMBO Rep* **10**, 614-21 (2009).
51. Toyofuku, T. et al. FARP2 triggers signals for Sema3A-mediated axonal repulsion. *Nat Neurosci* **8**, 1712-9 (2005).
52. Driessens, M.H., Olivo, C., Nagata, K., Inagaki, M. & Collard, J.G. B plexins activate Rho through PDZ-RhoGEF. *FEBS Lett* **529**, 168-72 (2002).
53. Barberis, D. et al. p190 Rho-GTPase activating protein associates with plexins and it is required for semaphorin signalling. *J Cell Sci* **118**, 4689-700 (2005).
54. Swiercz, J.M., Worzfeld, T. & Offermanns, S. ErbB-2 and met reciprocally regulate cellular signaling via plexin-B1. *J Biol Chem* **283**, 1893-901 (2008).
55. Grunicke, H. Signal transduction inhibitors. *Eur J Cancer* **31A**, 833-4 (1995).
56. Chung, H.H., Benson, D.R. & Schultz, P.G. Probing the structure and mechanism of Ras protein with an expanded genetic code. *Science* **259**, 806-9 (1993).
57. Etienne-Manneville, S. & Hall, A. Rho GTPases in cell biology. *Nature* **420**, 629-35 (2002).
58. Vetter, I.R. & Wittinghofer, A. The guanine nucleotide-binding switch in three dimensions. *Science* **294**, 1299-304 (2001).

59. Wittinghofer, A. & Vetter, I.R. Structure-function relationships of the g domain, a canonical switch motif. *Annu Rev Biochem* **80**, 943-71.
60. Valencia, A., Chardin, P., Wittinghofer, A. & Sander, C. The ras protein family: evolutionary tree and role of conserved amino acids. *Biochemistry* **30**, 4637-48 (1991).
61. Cerione, R.A. & Zheng, Y. The Dbl family of oncogenes. *Curr Opin Cell Biol* **8**, 216-22 (1996).
62. Bishop, A.L. & Hall, A. Rho GTPases and their effector proteins. *Biochem J* **348 Pt 2**, 241-55 (2000).
63. Lamarche, N. & Hall, A. GAPs for rho-related GTPases. *Trends Genet* **10**, 436-40 (1994).
64. Wittinghofer, A. Signaling mechanistics: aluminum fluoride for molecule of the year. *Curr Biol* **7**, R682-5 (1997).
65. Scheffzek, K. et al. The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* **277**, 333-8 (1997).
66. Scheffzek, K., Lautwein, A., Kabsch, W., Ahmadian, M.R. & Wittinghofer, A. Crystal structure of the GTPase-activating domain of human p120GAP and implications for the interaction with Ras. *Nature* **384**, 591-6 (1996).
67. Antipenko, A. et al. Structure of the semaphorin-3A receptor binding module. *Neuron* **39**, 589-98 (2003).
68. Love, C.A. et al. The ligand-binding face of the semaphorins revealed by the high-resolution crystal structure of SEMA4D. *Nat Struct Biol* **10**, 843-8 (2003).
69. Muller, K.M., Arndt, K.M. & Pluckthun, A. Model and simulation of multivalent binding to fixed ligands. *Anal Biochem* **261**, 149-58 (1998).
70. Monnier, P.P. et al. RGM is a repulsive guidance molecule for retinal axons. *Nature* **419**, 392-5 (2002).
71. Severyn, C.J., Shinde, U. & Rotwein, P. Molecular biology, genetics and biochemistry of the repulsive guidance molecule family. *Biochem J* **422**, 393-403 (2009).
72. Rajagopalan, S. et al. Neogenin mediates the action of repulsive guidance molecule. *Nat Cell Biol* **6**, 756-62 (2004).
73. Camus, L.M. & Lambert, L.A. Molecular evolution of hemojuvelin and the repulsive guidance molecule family. *J Mol Evol* **65**, 68-81 (2007).
74. Yang, F., West, A.P., Jr., 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).
75. Niederkofler, 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).
76. 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).
77. 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).

78. 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).
79. Hata, K., Kaibuchi, K., Inagaki, S. & Yamashita, T. Unc5B associates with LARG to mediate the action of repulsive guidance molecule. *J Cell Biol* **184**, 737-50 (2009).
80. 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).
81. 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).
82. Corradini, E., Babitt, J.L. & Lin, H.Y. The RGM/DRAGON family of BMP co-receptors. *Cytokine Growth Factor Rev* **20**, 389-98 (2009).
83. 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).
84. Xia, Y. et al. Dragon enhances BMP signaling and increases transepithelial resistance in kidney epithelial cells. *J Am Soc Nephrol* **21**, 666-77.
85. Xia, Y. et al. Dragon (repulsive guidance molecule b) inhibits IL-6 expression in macrophages. *J Immunol* **186**, 1369-76.
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.
87. Kuninger, D., Kuzmickas, R., Peng, B., Pintar, J.E. & Rotwein, P. Gene discovery by microarray: identification of novel genes induced during growth factor-mediated muscle cell survival and differentiation. *Genomics* **84**, 876-89 (2004).
88. Papanikolaou, G. et al. Mutations in HFE2 cause iron overload in chromosome 1q-linked juvenile hemochromatosis. *Nat Genet* **36**, 77-82 (2004).
89. Schmidtmer, J. & Engelkamp, D. Isolation and expression pattern of three mouse homologues of chick Rgm. *Gene Expr Patterns* **4**, 105-10 (2004).
90. Kuninger, D., Kuns-Hashimoto, R., Nili, M. & Rotwein, P. Pro-protein convertases control the maturation and processing of the iron-regulatory protein, RGMc/hemojuvelin. *BMC Biochem* **9**, 9 (2008).
91. Kuninger, D., Kuns-Hashimoto, R., Kuzmickas, R. & Rotwein, P. Complex biosynthesis of the muscle-enriched iron regulator RGMc. *J Cell Sci* **119**, 3273-83 (2006).
92. Andriopoulos, B., Jr. et al. BMP6 is a key endogenous regulator of hepcidin expression and iron metabolism. *Nat Genet* **41**, 482-7 (2009).
93. Lee, D.H. et al. Neogenin inhibits HJV secretion and regulates BMP-induced hepcidin expression and iron homeostasis. *Blood* **115**, 3136-45.
94. Tessier-Lavigne, M. & Goodman, C.S. The molecular biology of axon guidance. *Science* **274**, 1123-33 (1996).
95. 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).
96. Park, K.W. et al. The axonal attractant Netrin-1 is an angiogenic factor. *Proc Natl Acad Sci U S A* **101**, 16210-5 (2004).

97. Srinivasan, K., Strickland, P., Valdes, A., Shin, G.C. & Hinck, L. Netrin-1/neogenin interaction stabilizes multipotent progenitor cap cells during mammary gland morphogenesis. *Dev Cell* **4**, 371-82 (2003).
98. 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).
99. Kang, J.S. et al. Netrins and neogenin promote myotube formation. *J Cell Biol* **167**, 493-504 (2004).
100. Ren, X.R. et al. Focal adhesion kinase in netrin-1 signaling. *Nat Neurosci* **7**, 1204-12 (2004).
101. Hata, K. et al. RGMa inhibition promotes axonal growth and recovery after spinal cord injury. *J Cell Biol* **173**, 47-58 (2006).
102. Kubo, T., Hata, K., Yamaguchi, A. & Yamashita, T. Rho-ROCK inhibitors as emerging strategies to promote nerve regeneration. *Curr Pharm Des* **13**, 2493-9 (2007).
103. Moore, S.W., Tessier-Lavigne, M. & Kennedy, T.E. Netrins and their receptors. *Adv Exp Med Biol* **621**, 17-31 (2007).
104. De Vries, M. & Cooper, H.M. Emerging roles for neogenin and its ligands in CNS development. *J Neurochem* **106**, 1483-92 (2008).
105. Katagiri, T., Suda, T., Miyazono, K. *The TGF-β Family*, (Cold Spring Harbor Laboratory Press, New York, 2008).
106. Chen, D., Zhao, M. & Mundy, G.R. Bone morphogenetic proteins. *Growth Factors* **22**, 233-41 (2004).
107. Shi, Y. & Massague, J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685-700 (2003).
108. Heldin, C.H., Miyazono, K. & ten Dijke, P. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature* **390**, 465-71 (1997).
109. Hoodless, P.A. et al. MADR1, a MAD-related protein that functions in BMP2 signaling pathways. *Cell* **85**, 489-500 (1996).
110. Nakao, A. et al. TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4. *EMBO J* **16**, 5353-62 (1997).
111. Zhou, Z. et al. Neogenin regulation of BMP-induced canonical Smad signaling and endochondral bone formation. *Dev Cell* **19**, 90-102.
112. 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).
113. 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).
114. 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.
115. Chang, V.T. et al. Glycoprotein structural genomics: solving the glycosylation problem. *Structure* **15**, 267-73 (2007).

116. 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).
117. Mayo, C.J. et al. Benefits of automated crystallization plate tracking, imaging, and analysis. *Structure* **13**, 175-82 (2005).
118. Leslie, A.G.W. Recent changes to the Mosflm package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**(1992).
119. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* **276**, 307-326 (1997).
120. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3 (1994).
121. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol Biol* **364**, 215-30 (2007).
122. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61**, 458-64 (2005).
123. Perrakis, A., Morris, R. & Lamzin, V.S. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**, 458-63 (1999).
124. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
125. Blanc, E. et al. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr* **60**, 2210-21 (2004).
126. Adams, P.D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-54 (2002).
127. Davis, I.W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-83 (2007).
128. Painter, J. & Merritt, E.A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62**, 439-50 (2006).
129. Turner, L.J. & Hall, A. Plexin-induced collapse assay in COS cells. *Methods Enzymol* **406**, 665-76 (2006).
130. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* **78**, 1606-19 (2000).
131. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881-90 (1988).
132. Gouet, P., Courcelle, E., Stuart, D.I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-8 (1999).
133. DeLano, W.L. The PyMol molecular graphics system. *DeLano Scientific, San Carlos, CA, USA* (2002).
134. Siryaporn, A. & Goulian, M. Cross-talk suppression between the CpxA-CpxR and EnvZ-OmpR two-component systems in E. coli. *Mol Microbiol* **70**, 494-506 (2008).
135. Laub, M.T. & Goulian, M. Specificity in two-component signal transduction pathways. *Annu Rev Genet* **41**, 121-45 (2007).

136. Bourret, R.B., Hess, J.F. & Simon, M.I. Conserved aspartate residues and phosphorylation in signal transduction by the chemotaxis protein CheY. *Proc Natl Acad Sci U S A* **87**, 41-5 (1990).
137. West, A.H. & Stock, A.M. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* **26**, 369-76 (2001).
138. Porter, S.L., Wadhams, G.H. & Armitage, J.P. Rhodobacter sphaeroides: complexity in chemotactic signalling. *Trends Microbiol* **16**, 251-60 (2008).
139. Wadhams, G.H. & Armitage, J.P. Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol* **5**, 1024-37 (2004).
140. Martin, A.C. et al. CheR- and CheB-dependent chemosensory adaptation system of Rhodobacter sphaeroides. *J Bacteriol* **183**, 7135-44 (2001).
141. Porter, S.L. & Armitage, J.P. Phosphotransfer in Rhodobacter sphaeroides chemotaxis. *J Mol Biol* **324**, 35-45 (2002).
142. Wadhams, G.H. et al. TlpC, a novel chemotaxis protein in Rhodobacter sphaeroides, localizes to a discrete region in the cytoplasm. *Mol Microbiol* **46**, 1211-21 (2002).
143. Wadhams, G.H., Warren, A.V., Martin, A.C. & Armitage, J.P. Targeting of two signal transduction pathways to different regions of the bacterial cell. *Mol Microbiol* **50**, 763-70 (2003).
144. Porter, S.L., Warren, A.V., Martin, A.C. & Armitage, J.P. The third chemotaxis locus of Rhodobacter sphaeroides is essential for chemotaxis. *Mol Microbiol* **46**, 1081-94 (2002).
145. Schneider, T.R. & Sheldrick, G.M. Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58**, 1772-9 (2002).
146. Langer, G., Cohen, S.X., Lamzin, V.S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**, 1171-9 (2008).
147. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-55 (1997).
148. Volz, K. & Matsumura, P. Crystal structure of Escherichia coli CheY refined at 1.7-A resolution. *J Biol Chem* **266**, 15511-9 (1991).
149. Stock, A.M. et al. Structure of the Mg(2+)-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. *Biochemistry* **32**, 13375-80 (1993).
150. Zhou, H. & Dahlquist, F.W. Phosphotransfer site of the chemotaxis-specific protein kinase CheA as revealed by NMR. *Biochemistry* **36**, 699-710 (1997).
151. Mourey, L. et al. Crystal structure of the CheA histidine phosphotransfer domain that mediates response regulator phosphorylation in bacterial chemotaxis. *J Biol Chem* **276**, 31074-82 (2001).
152. Quezada, C.M., Gradinaru, C., Simon, M.I., Bilwes, A.M. & Crane, B.R. Helical shifts generate two distinct conformers in the atomic resolution structure of the CheA phosphotransferase domain from Thermotoga maritima. *J Mol Biol* **341**, 1283-94 (2004).
153. Nooren, I.M. & Thornton, J.M. Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol* **325**, 991-1018 (2003).

154. Skerker, J.M. et al. Rewiring the specificity of two-component signal transduction systems. *Cell* **133**, 1043-54 (2008).
155. Szurmant, H. et al. Co-evolving motions at protein-protein interfaces of two-component signaling systems identified by covariance analysis. *Biochemistry* **47**, 7782-4 (2008).
156. Weigt, M., White, R.A., Szurmant, H., Hoch, J.A. & Hwa, T. Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci U S A* **106**, 67-72 (2009).
157. Varughese, K.I., Tsigelny, I. & Zhao, H. The crystal structure of beryllofluoride Spo0F in complex with the phosphotransferase Spo0B represents a phosphotransfer pretransition state. *J Bacteriol* **188**, 4970-7 (2006).
158. Zapf, J., Sen, U., Madhusudan, Hoch, J.A. & Varughese, K.I. A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. *Structure* **8**, 851-62 (2000).
159. Xu, Q., Porter, S.W. & West, A.H. The yeast YPD1/SLN1 complex: insights into molecular recognition in two-component signaling systems. *Structure* **11**, 1569-81 (2003).
160. Zhao, X., Copeland, D.M., Soares, A.S. & West, A.H. Crystal structure of a complex between the phosphorelay protein YPD1 and the response regulator domain of SLN1 bound to a phosphoryl analog. *J Mol Biol* **375**, 1141-51 (2008).
161. Thomas, S.A., Brewster, J.A. & Bourret, R.B. Two variable active site residues modulate response regulator phosphoryl group stability. *Mol Microbiol* **69**, 453-65 (2008).
162. Galperin, M.Y. Diversity of structure and function of response regulator output domains. *Curr Opin Microbiol* **13**, 150-9.
163. Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. & Iyer, L.M. The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* **29**, 231-62 (2005).
164. Baikalov, I. et al. Structure of the Escherichia coli response regulator NarL. *Biochemistry* **35**, 11053-61 (1996).
165. Buckler, D.R., Zhou, Y. & Stock, A.M. Evidence of intradomain and interdomain flexibility in an OmpR/PhoB homolog from Thermotoga maritima. *Structure* **10**, 153-64 (2002).
166. Hengge, R. Principles of c-di-GMP signalling in bacteria. *Nat Rev Microbiol* **7**, 263-73 (2009).
167. Francez-Charlot, A. et al. Sigma factor mimicry involved in regulation of general stress response. *Proc Natl Acad Sci U S A* **106**, 3467-72 (2009).
168. Bourret, R.B. Receiver domain structure and function in response regulator proteins. *Curr Opin Microbiol* **13**, 142-9.
169. Gao, R. & Stock, A.M. Molecular strategies for phosphorylation-mediated regulation of response regulator activity. *Curr Opin Microbiol* **13**, 160-7.
170. Lee, S.Y. et al. Crystal structure of an activated response regulator bound to its target. *Nat Struct Biol* **8**, 52-6 (2001).

171. Bren, A. & Eisenbach, M. The N terminus of the flagellar switch protein, FliM, is the binding domain for the chemotactic response regulator, CheY. *J Mol Biol* **278**, 507-14 (1998).
172. Kern, D. et al. Structure of a transiently phosphorylated switch in bacterial signal transduction. *Nature* **402**, 894-8 (1999).
173. Lei, M. et al. Segmented transition pathway of the signaling protein nitrogen regulatory protein C. *J Mol Biol* **392**, 823-36 (2009).
174. Gardino, A.K. et al. Transient non-native hydrogen bonds promote activation of a signaling protein. *Cell* **139**, 1109-18 (2009).
175. Lee, S.Y. et al. Crystal structure of activated CheY. Comparison with other activated receiver domains. *J Biol Chem* **276**, 16425-31 (2001).
176. Claude, J.B., Suhre, K., Notredame, C., Claverie, J.M. & Abergel, C. CaspR: a web server for automated molecular replacement using homology modelling. *Nucleic Acids Res* **32**, W606-9 (2004).
177. Notredame, C., Higgins, D.G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-17 (2000).
178. Sali, A. & Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815 (1993).
179. Navaza, J. Implementation of molecular replacement in AMoRe. *Acta Crystallogr D Biol Crystallogr* **57**, 1367-72 (2001).
180. Brunger, A.T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-21 (1998).
181. Birck, C., Chen, Y., Hulett, F.M. & Samama, J.P. The crystal structure of the phosphorylation domain in PhoP reveals a functional tandem association mediated by an asymmetric interface. *J Bacteriol* **185**, 254-61 (2003).
182. Volz, K. Structural conservation in the CheY superfamily. *Biochemistry* **32**, 11741-53 (1993).
183. Wuthrich, K. *NMR in biological research: Peptides and proteins*, 379 (American Elsevier Pub. Co, 1976).
184. Riepl, H., Scharf, B., Schmitt, R., Kalbitzer, H.R. & Maurer, T. Solution structures of the inactive and BeF3-activated response regulator CheY2. *J Mol Biol* **338**, 287-97 (2004).
185. Dyer, C.M. & Dahlquist, F.W. Switched or not?: the structure of unphosphorylated CheY bound to the N terminus of FliM. *J Bacteriol* **188**, 7354-63 (2006).
186. Tran, T.S., Kolodkin, A.L. & Bharadwaj, R. Semaphorin regulation of cellular morphology. *Annu Rev Cell Dev Biol* **23**, 263-92 (2007).
187. Balakrishnan, A. et al. Molecular profiling of the "plexinome" in melanoma and pancreatic cancer. *Hum Mutat* **30**, 1167-74 (2009).
188. Wong, O.G. et al. Plexin-B1 mutations in prostate cancer. *Proc Natl Acad Sci U S A* **104**, 19040-5 (2007).
189. Tamagnone, L. et al. Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. *Cell* **99**, 71-80 (1999).

190. Hu, H., Marton, T.F. & Goodman, C.S. Plexin B mediates axon guidance in Drosophila by simultaneously inhibiting active Rac and enhancing RhoA signaling. *Neuron* **32**, 39-51 (2001).
191. Bos, J.L., Rehmann, H. & Wittinghofer, A. GEFs and GAPs: critical elements in the control of small G proteins. *Cell* **129**, 865-77 (2007).
192. Oinuma, I., Katoh, H. & Negishi, M. Semaphorin 4D/Plexin-B1-mediated R-Ras GAP activity inhibits cell migration by regulating beta(1) integrin activity. *J Cell Biol* **173**, 601-13 (2006).
193. Driessens, M.H. et al. Plexin-B semaphorin receptors interact directly with active Rac and regulate the actin cytoskeleton by activating Rho. *Curr Biol* **11**, 339-44 (2001).
194. Tong, Y. et al. Binding of Rac1, Rnd1, and RhoD to a novel Rho GTPase interaction motif destabilizes dimerization of the plexin-B1 effector domain. *J Biol Chem* **282**, 37215-24 (2007).
195. Vikis, H.G., Li, W., He, Z. & Guan, K.L. The semaphorin receptor plexin-B1 specifically interacts with active Rac in a ligand-dependent manner. *Proc Natl Acad Sci U S A* **97**, 12457-62 (2000).
196. Negishi, M. & Katoh, H. Rho family GTPases as key regulators for neuronal network formation. *J Biochem* **132**, 157-66 (2002).
197. Turner, L.J., Nicholls, S. & Hall, A. The activity of the plexin-A1 receptor is regulated by Rac. *J Biol Chem* **279**, 33199-205 (2004).
198. Vikis, H.G., Li, W. & Guan, K.L. The plexin-B1/Rac interaction inhibits PAK activation and enhances Sema4D ligand binding. *Genes Dev* **16**, 836-45 (2002).
199. Oinuma, I., Katoh, H. & Negishi, M. Molecular dissection of the semaphorin 4D receptor plexin-B1-stimulated R-Ras GTPase-activating protein activity and neurite remodeling in hippocampal neurons. *J Neurosci* **24**, 11473-80 (2004).
200. Tong, Y., Hota, P.K., Hamaneh, M.B. & Buck, M. Insights into oncogenic mutations of plexin-B1 based on the solution structure of the Rho GTPase binding domain. *Structure* **16**, 246-58 (2008).
201. Tong, Y., Hughes, D., Placanica, L. & Buck, M. When monomers are preferred: a strategy for the identification and disruption of weakly oligomerized proteins. *Structure* **13**, 7-15 (2005).
202. Best, A., Ahmed, S., Kozma, R. & Lim, L. The Ras-related GTPase Rac1 binds tubulin. *J Biol Chem* **271**, 3756-62 (1996).
203. Hirshberg, M., Stockley, R.W., Dodson, G. & Webb, M.R. The crystal structure of human rac1, a member of the rho-family complexed with a GTP analogue. *Nat Struct Biol* **4**, 147-52 (1997).
204. Stuart, D.I., Levine, M., Muirhead, H. & Stammers, D.K. Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 A. *J Mol Biol* **134**, 109-42 (1979).
205. Hakoshima, T., Shimizu, T. & Maesaki, R. Structural basis of the Rho GTPase signaling. *J Biochem* **134**, 327-31 (2003).
206. Riento, K., Villalonga, P., Garg, R. & Ridley, A. Function and regulation of RhoE. *Biochem Soc Trans* **33**, 649-51 (2005).
207. Grasberger, B., Minton, A.P., DeLisi, C. & Metzger, H. Interaction between proteins localized in membranes. *Proc Natl Acad Sci U S A* **83**, 6258-62 (1986).

208. Takahashi, T. & Strittmatter, S.M. Plexina1 autoinhibition by the plexin sema domain. *Neuron* **29**, 429-39 (2001).
209. 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).
210. 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.
211. Muramatsu, R. et al. RGMa modulates T cell responses and is involved in autoimmune encephalomyelitis. *Nat Med* **17**, 488-94.
212. Keino-Masu, K. et al. Deleted in Colorectal Cancer (DCC) encodes a netrin receptor. *Cell* **87**, 175-85 (1996).
213. Miyazawa, K., Shinozaki, M., Hara, T., Furuya, T. & Miyazono, K. Two major Smad pathways in TGF-beta superfamily signalling. *Genes Cells* **7**, 1191-204 (2002).
214. Nemeth, E. Iron regulation and erythropoiesis. *Curr Opin Hematol* **15**, 169-75 (2008).
215. Lanzara, C. et al. Spectrum of hemojuvelin gene mutations in 1q-linked juvenile hemochromatosis. *Blood* **103**, 4317-21 (2004).
216. Babitt, J.L. et al. Bone morphogenetic protein signaling by hemojuvelin regulates hepcidin expression. *Nat Genet* **38**, 531-9 (2006).
217. 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).
218. Lin, L. et al. Soluble hemojuvelin is released by proprotein convertase-mediated cleavage at a conserved polybasic RNRR site. *Blood Cells Mol Dis* **40**, 122-31 (2008).
219. Yang, F., West, A.P., Jr. & Bjorkman, P.J. Crystal structure of a hemojuvelin-binding fragment of neogenin at 1.8A. *J Struct Biol* **174**, 239-44.
220. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-68 (2004).
221. Ye, Y. & Godzik, A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* **19 Suppl 2**, ii246-55 (2003).
222. Guerler, A. & Knapp, E.W. Novel protein folds and their nonsequential structural analogs. *Protein Sci* **17**, 1374-82 (2008).
223. Landon. Cleavage at aspartyl-prolyl bonds. *Methods Enzymol* **47**, 145-9 (1977).
224. 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).
225. 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).
226. Piszkiewicz, D., Landon, M. & Smith, E.L. Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations. *Biochem Biophys Res Commun* **40**, 1173-8 (1970).
227. Demaurex, N., Furuya, W., D'Souza, S., Bonifacino, J.S. & Grinstein, S. Mechanism of acidification of the trans-Golgi network (TGN). In situ measurements of pH using retrieval of TGN38 and furin from the cell surface. *J Biol Chem* **273**, 2044-51 (1998).

228. Miesenbock, G., De Angelis, D.A. & Rothman, J.E. Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* **394**, 192-5 (1998).
229. Lee, P.L., Barton, J.C., Brandhagen, D. & Beutler, E. Hemojuvelin (HJV) mutations in persons of European, African-American and Asian ancestry with adult onset haemochromatosis. *Br J Haematol* **127**, 224-9 (2004).
230. Lee, P.L., Beutler, E., Rao, S.V. & Barton, J.C. Genetic abnormalities and juvenile hemochromatosis: mutations of the HJV gene encoding hemojuvelin. *Blood* **103**, 4669-71 (2004).
231. Sadler, J.E. Biochemistry and genetics of von Willebrand factor. *Annu Rev Biochem* **67**, 395-424 (1998).
232. Meiniel, O. & Meiniel, A. The complex multidomain organization of SCO-spondin protein is highly conserved in mammals. *Brain Res Rev* **53**, 321-7 (2007).
233. Kuns-Hashimoto, R., Kuninger, D., Nili, M. & Rotwein, P. Selective binding of RGMc/hemojuvelin, a key protein in systemic iron metabolism, to BMP-2 and neogenin. *Am J Physiol Cell Physiol* **294**, C994-C1003 (2008).
234. Hauburger, A. et al. The pro-form of BMP-2 interferes with BMP-2 signalling by competing with BMP-2 for IA receptor binding. *FEBS J* **276**, 6386-98 (2009).
235. Sako, D. et al. Characterization of the ligand binding functionality of the extracellular domain of activin receptor type IIb. *J Biol Chem* **285**, 21037-48.
236. Moser, M. et al. BMPER, a novel endothelial cell precursor-derived protein, antagonizes bone morphogenetic protein signaling and endothelial cell differentiation. *Mol Cell Biol* **23**, 5664-79 (2003).
237. Porter, S.L., Wadhams, G.H. & Armitage J.P. Signal processing in complex chemotactic pathways. *Nat Rev Microbiol* **9**, 153-65 (2011).
238. Laskowski, R.A. PDBsum new things. *Nucleic Acids Res* **37**, 355-9 (2008).