The Cell Surface Organisation of the Notch-1 Receptor



Philip Weisshuhn

Department of Biochemistry

Linacre College

A thesis submitted for the degree of Doctor of Philosophy at the

University of Oxford

Abstract

The Cell Surface Organisation of the Notch-1 Receptor

Philip Weisshuhn, Linacre College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy, Trinity Term 2012

The Notch receptor family plays a key role in development and disease. In cancer, Notch can act either as an oncogene or as a tumour suppressor, and possibly as a cancer stem-cell factor. Whereas most research has focused on downstream signalling events, little is known about the cell surface organisation of Notch and its ligands. The extracellular part of Notch consists mainly of 36 epidermal growth factor-like domains (EGF-domains), many of which bind calcium. Studies have shown that tandem repeats of calcium-binding EGF domains form a rigid linear arrangement; however, the lack of calcium binding in EGF6, EGF10 and EGF22 led to the hypothesis that these might be sites of flexibility.

This thesis addresses the effect of these domains on the organisation of the extracellular region of Notch and provides further insight into the calcium-binding properties of Notch. NMR residual dipolar coupling (RDC) measurements of these regions are presented, together with the X-ray crystallographic data obtained in collaboration.

The crystal structure of the human Notch-1 construct EGF4-7 shows a tilt angle of 90° at the EGF5-6 interface which is much larger than the tilt angles of 10-20° observed for the EGF11-13 crystal structure. RDC measurements demonstrated an angle of ~70° in solution. The crystal structures of EGF21-23 and EGF20-23 showed a rod-shaped interface for the EGF21-22 domain, in which a *cis*-proline forms the packing interaction to a tyrosine at the β-turn in the major β-sheet of EGF22. These two interfaces are novel and demonstrate the possibility of interface formation without Ca^{2+}. Crystallisation was unsuccessful for the EGF8-11 construct. However, RDC measurements indicate interdomain motion between EGF9 and EGF10 demonstrating a flexible interface.

These data establish new information on the structural organisation and calcium-binding properties of the extracellular region of Notch and identify flexible and rigid interfaces within multiple tandem repeats of EGF domains. This information will be invaluable in constructing models of Notch-ligand complexes for testing in future functional experiments.

Acknowledgements

The majority of the work presented in this thesis is based on techniques and knowledge which were established and acquired respectively in the laboratories of Professor Christina Redfield and Professor Penny Handford, who guided me through my thesis as my supervisors.

Professor Christina Redfield, with her hands-on approach on complicated NMR techniques, continuously supported me during the planning and setting up of all kinds of NMR and other experiments performed throughout this work and the subsequent analysis. The training in the field of biophysical chemistry I received from Professor Redfield provided me with the necessary theoretic framework to tackle this thesis. I am very thankful that in her guidance, Professor Redfield was always optimistic; this worked as a great source of mental energy.

Professor Handford had a very close eye on my overall performance, as a proper scientific teacher with advice at the right place; comments such as „don't rush“ were all spot-on and helped me shaping and developing at the critical steps during my training. Professor Handford's teachings on the scientific method, e.g., the necessity of controls, are legendary and taught me to scrutinise my own experiments to the point of full confidence. Professor Handford's explanations initiated the cultivation of awareness of the given situation and the necessity to develop strategies to get into control over the experiments and to plan for in cases when high-risk experiments could fail.

My research group worked as a team and many scientists hence contributed in shaping my thesis: Drs Sarah Iqbal, Juliana Callaghan, David Yadin, Lukas Stelzl, Matt Damon, Ian Robertson, Paul Taylor, Sophie Gilbert, Shaoyan Liang, and Ralf Schneider. In particular, Dr Sacha Jensen was extremely supportive at all stages of my thesis, dealing with matters ranging from a pure technical point of view, to experiment design and to questions of friendship such as teaching me piano.

I would further like to thank Drs David Staunton, Jonathan Boyd and Nick Soffe for providing me with technical expertise and Martin Ackland for computer support. I would like to thank Dr Devon Sheppard and Professor Susan Lea from the Sir William Dunn School of Pathology for crystallizing my constructs. And, I would also like to thank Dr Jason Schnell, Dr Ioannis Vakonakis and Professor Iain Campbell for further advice.

During my four years at Oxford, a close circle of friends at Linacre College was formed, reflecting the splendid system at Oxford which gives each student a working place in an academically outstanding research department, but also a family-like home in one of the different colleges. In the first year of my DPhil, the tiny kitchen on the first floor of the Abraham's building became my favourite spot to arrive home after a long and exhausting day in the laboratory. In the Abraham kitchen, I would always find a friendship-enriched atmosphere and being offered comfort food according to recipes from all over the world: Siriol McAvoy (Wales), Natalie Christopher (Cyprus), Oliver Brain (South Africa), Karine Peloffy (French-Canada), Nicole Novak (Iowa) and Ruben Kubiak (Germany). In the second year, I was further delighted with Kantonese meals by Jessie Fu, Korean meals by Sophia Lee and Sushi by Yukihiro Nomura. In my third year, in which I often did not leave the laboratory before 11pm, I would cherish Adwoa Darko-Mensah from Ghana that would always prepare an extra-portion of delicious Jollof-rice, rich in chilly to spark the fading energy levels of the third year. In my fourth year, I became friends with the most kind-hearted British students Niel Bowerman and Eleanor James, and lived up to the end of my DPhil in a circle of friends with Julia Yang, Oliver Winkenbach, Laura Keating, Melanie Sindelar and Yoonyoung Min. A further thank you goes to my Linacre friends Tommy Kho, Venus Tsang, Chelsea Robles for various cooking sessions. Linacre College further emphasised the need for physical wellbeing greatly. The gym was city-wide known for its quality and I was taught how to lift, pull and push barbells in and from all directions by Jörn Klinger, Erik Makinen and Patrick Lutaya. I would like to thank the people

involved in the rowing and tennis club. I am very thankful to Lt Col Patrick Heffernan, the back then Bursar of Linacre College with whom I worked together at Linacre. After serving for many years as a British soldier in Germany, his friendly and caring way gave me the best example of how British-German friendship should unfold. Last, but not least, a thank you to Professor Nicole Zitzmann, who was my college advisor at Linacre and with which I had fruitful and very pleasant discussions and reunions and who always insisted to talk on important matters which were not necessarily science based.

All this work would not have been possible without the people described. However, my most sincere thank you goes to Cancer Research UK for collecting donations and providing me with full funding and an additional research grant to be able to contribute to a world without cancer.

I dedicate this thesis to all the people suffering from cancer and other diseases that are difficult to treat. May they never give up hope and may the scientific community never stop investigating better treatment options.

Abbreviations

ANK	Ankyrin
BTD	β-trefoil domain
CADASIL	Cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy
CSL	CBF1/RBP-Jκ/Suppressor of Hairless/LAG-1
cbEGF	Ca^{2+} binding EGF
DSL	Delta/Serrate/Lag-2
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor-like
FPLC	Fast protein liquid chromatography
HD	Heterodimerization domain
hN-1	human Notch-1
HPLC	High-performance liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K_d	Dissociation constant
kDa	Kilodalton
LNR	LIN-12/Notch repeat
MAML-1	Mastermind-like protein 1
mAU	Milliabsorbance units
MHz	Megahertz
MNNL	Module at the N-terminus of Notch ligands
mN-1	Mouse Notch-1
NICD	Notch intracellular domain
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NRR	Negative regulatory region
PAGE	Polyacrylamide gel electrophoresis
PDB-ID	Protein database identifier
PEST	Peptide sequence rich in proline (P), glutamic acid (E), serine (S), and threonine (T)
Ppm	Parts per million
RAM	Recombination binding protein-J associated molecule
RDC	Residual dipolar coupling
Rp	Reverse-phase
Rpm	Revolutions per minute
RT	Room temperature
SDS	Sodium Dodecyl Sulfate
TACE	TNF-α converting enzyme
T-ALL	T-cell acute lymphoblastic leukaemia
Tan1	Translocation-associated Notch homolog-1
TB	Transforming growth factor β-binding protein-like
TFA	Trifluoroacetic acid
TM	Transmembrane

Contents

Title page	i
Abstract	ii
Acknowledgements	iii
Abbreviations	vi
Contents	vii
1 Introduction	1
1.1 Biology of Notch	2
1.1.1 Overview	2
1.1.2 Canonical Notch signalling – a timeline of discovery	3
1.1.3 Notch - three modes of action	5
1.2 The structural basis for Notch signalling	6
1.2.1 The core pathway and key players	6
1.2.2 Structure of the ligand binding region	7
1.2.3 *Cis*-interactions	11
1.2.4 Receptor activation	12
1.2.5 Transcriptional regulation	14
1.3 Notch in disease	15
1.3.1 Oncogenic functions	16
1.3.2 Tumour suppressor functions	17
1.3.3 Cancer therapy	17
1.3.4 Alagille Syndrome	19
1.4 The cell surface organisation of Notch-1	19
1.4.1 The structure of EGF domains	20
1.4.2 Calcium binding EGF (cbEGF) domains	22
1.4.3 Posttranslational modifications	23
1.4.4 cbEGF domain interfaces	27
1.4.5 Interfaces without Ca^{2+}	32
1.4.6 Summary and aim of thesis	33
2 Materials and methods	35
2.1 Amplification and cloning of DNA	36
2.2 Expression and purification of protein	38
2.3 *In-vitro* refolding and further purification	39

2.4 Determination of protein concentration 39
2.5 SDS-PAGE 40
2.6 Ca^{2+} chelator competition assays 40
2.7 NMR experiments 41
2.8 ^{1}H NMR Ca^{2+} titration studies 41
2.9 ^{1}H-^{15}N and triple labelled NMR experiments 42
2.10 RDC experiments 42
2.11 Calculation of torsion angles 44
2.12 Crystallisation 44
2.13 Cell aggregation assay 45
3 Characterisation of protein constructs 47
3.1 Introduction 48
3.2 Expression of protein 48
3.3 Refolding and purification of hN-1_{7-9} 48
3.4 FPLC ion-exchange chromatography and further purification 49
3.5 SDS PAGE and Mass spectrometry 50
3.6 1D ^{1}H NMR Analysis 52
4 Ca^{2+} dissociation constants 59
4.1 Introduction 60
4.2 Determination of Ca^{2+} affinities 64
4.2.1 Binding models and affinity 64
4.2.2 NMR in measurement of low Ca^{2+} affinities 64
4.2.3 Chromophoric chelation to measure high Ca^{2+} affinities 66
4.3 Calcium binding in hN-1_{4-7} and hN-1_{5-7} 67
4.4 Calcium binding in hN-1_{9-11} and hN-1_{8-11} 71
4.5 Calcium binding in hN-1_{7-9} 76
4.6 Discussion 78
5 Assignments of chemical shifts 82
5.1 Introduction 83
5.2. Results 84
5.2.1 HSQC analysis 84
5.2.2 Overlapping constructs 86
5.2.3 3D NOESY/TOCSY spectroscopy 86
5.2.4 Triple resonance experiments 88
5.2.5 Further optimisation and the hN-1_{8-11} construct 90

5.3 Conclusion and extent of assignment 94
6 Structural analysis 99
6.1 Introduction 100
6.2 Structural analysis of hN-1_{11-13} 101
6.2.1 Ca^{2+} binding in hN-1_{11-13} 101
6.2.2 Phi/Psi torsion angle comparison for hN-1_{11-13} 104
6.3 Crystallography 106
6.4 Domain orientation 108
6.5 Structural analysis of hN-1_{4-7} 109
6.5.1 Overview of structures 109
6.5.2 The EGF5-EGF6 interface 115
6.5.2.1 NOESY analysis of the EGF5-6 interface 118
6.5.2.2 EGF5-EGF6 cell aggregation assay 120
6.5.3 Ca^{2+} binding of EGF5 and EGF7 123
6.5.4 The EGF4-EGF5 interface 124
6.5.5 The EGF6-EGF7 interface 126
6.6 Structural analysis of hN-1_{21-23} 127
6.6.1 The EGF21-EGF22 interface 127
6.6.2 The EGF22-EGF23 interface 129
6.7 Discussion 131
7 TALOS+ analysis 134
7.1 Introduction to TALOS+ and dynamics 135
7.2 TALOS+ torsion angle predictions 138
7.2.1 Predictions for individual domains in hN-1_{11-13} 138
7.2.2 Accuracy of torsion angle predictions in hN-1_{11-13} 140
7.2.3 Predictions for individual domains in hN-1_{4-7} 143
7.2.4 Accuracy of torsion angle predictions in hN-1_{4-7} 148
7.2.5 Predictions for individual domains in hN-1_{8-11} 150
7.3 Secondary structure analysis 155
7.4 Fast-timescale dynamics 157
7.5 Ca^{2+} backbone ligands 160
7.6 Discussion 162
8 Defining inter-domain orientations using residual dipolar couplings 164
8.1 Introduction 165
8.2 RDC measurements 167
8.2.1 Theory 167

8.2.2 RDC measurements 169
8.2.3 Domain orientation 172
8.3 Analysis of the hN-1_{11-13} domain architecture 175
8.3.1 RDC measurement in hN-1_{11-13} 175
8.3.2 Alignment tensors in individual domains 176
8.3.3 Domain orientation 180
8.3.4 Analysis of the hN-1_{11-13}12DG mutant 183
8.4 Analysis of the hN-1_{4-7} domain architecture 190
8.5 Analysis of the EGF7-8 domain architecture 197
8.6 Analysis of the hN-1_{8-11} domain architecture 202
8.7 Discussion 207
9 Discussion 210
9.1 Aim of thesis 211
9.2 Results 212
9.3 Outlook 215
9.4 Summary 217
10 References 218
Appendix 232

1 Introduction

1.1 Biology of Notch

1.1.1 Overview

Cell-cell communication is crucial for multicellular organisms; their development starting from a single cell and the cellular processes that govern the maintenance of tissue are highly orchestrated. In metazoa, a small number of core signal pathways evolved that iteratively control cell proliferation, numerous cell-fate specifications and cell death. The heterodimeric transmembrane adhesion receptor Notch belongs to these, and is unusual in that its ligands are usually transmembrane proteins; hence Notch signalling is limited to directly attached cells (Figure 1.1).

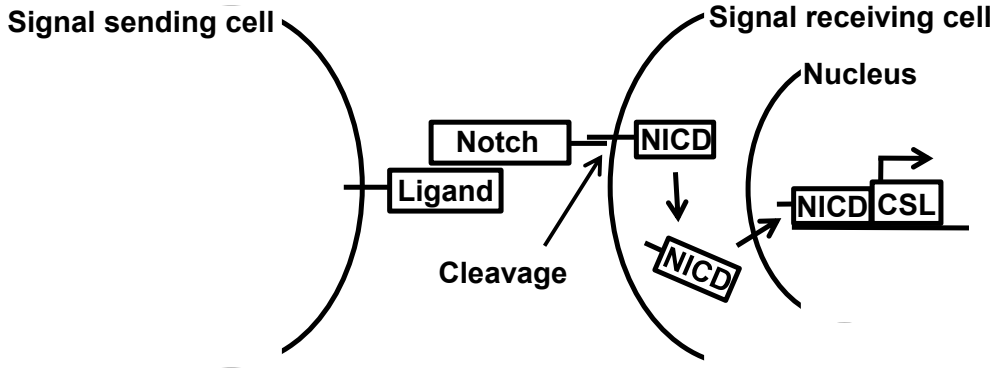


Figure 1.1 Initiation of the canonical Notch signalling pathway. Current models suggest that ligand binding in *trans* leads to cleavage events that result in the transport of the Notch intracellular domain (NICD) into the nucleus. Binding to the transcription factor family CSL activates gene expression.

The core Notch signalling pathways appears simple; Notch acts as both a receptor and a transcription factor. Layers of complexity are added by posttranslational modifications and the integration of Notch signals into a complex genetic circuitry. Aberrant Notch signalling is a direct cause in developmental or adult-onset genetic diseases such as Alagille and CADASIL syndrome, and Notch can act as either as an oncogene or a tumour suppressor.

Whereas the downstream signalling pathway has been studied in detail, little is known about the Notch-ligand binding complex and early events of activation. The extracellular region of Notch consists mainly of 36 EGF-like domains, and this thesis addresses their overall organisation.

1.1.2 Canonical Notch signalling – a timeline of discovery

Notch was first described by John S. Dexter in 1914 as a lethal phenotype that caused haploinsufficient notching in the wing of *Drosophila melanogaster* (Figure 1.2A) (Dexter, 1914). In 1916, this phenotype was mapped to the X-chromosome by Thomas Hunt Morgan (Morgan and Bridges, 1916).

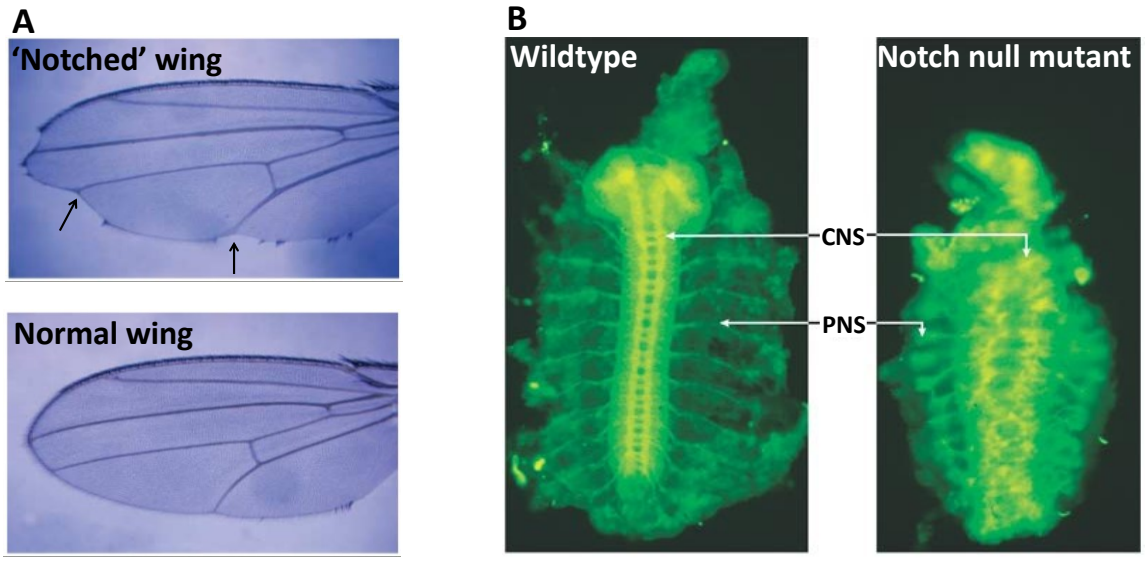


Figure 1.2 The Notch phenotype. (A) Haploinsufficiency causes notches (arrows) in the wing of *Drosophila melanogaster*. (B) Notch null mutations result in a disrupted segregation of the ectoderm with an overproduction of neuroblasts in both the central (CNS) and peripheral (PNS) nervous system (staining against neural tissue). Image adapted (Louvi and Artavanis-Tsakonas, 2006, Radtke and Raj, 2003).

Studies with null mutation embryos performed by Poulson revealed the full potential of the Notch signalling pathway – the ectoderm of embryos failed to segregate neural and epidermal lineages; only neuroblasts were produced at the expense of epidermal cells. The resulting hypertrophy of neural tissue is shown in Figure 1.2B (Poulson, 1940, Poulson, 1937).

Over the years, research on *Drosophila* emphasised the importance of Notch and its pleiotropic action, and Notch became a popular object of study. One of the first reviews on Notch started with the sentence "If one was asked to choose the single, most important genetic variation concerned with the expression of the genome during embryogenesis in *Drosophila melanogaster*, the answer would have to be the *Notch* locus." (Wright, 1970).

During the following years, a set of genes was described that resembled very much the neurogenic phenotype of Notch. These genes were classified as the "neurogenic locus" and later shown to encode for Notch core signalling pathway components. The Notch ligand Delta, for example, was observed in the third chromosome in *Drosophila* (Bridges and Morgan 1923, unpublished). Most of the other key players were identified in the next sixty years (Lehmann et al., 1983, Xu et al., 1990). The Notch gene sequence was first described in *Drosophila*, and encoded a 300-kDa single-pass transmembrane protein, rich in EGF-like domains on its extracellular region (Wharton et al., 1985). Subsequently two copies with homology to Notch were found in *C. elegans*, *glp-1* and *lin-12* (Austin and Kimble, 1987, Yochem et al., 1988, Yochem and Greenwald, 1989, Austin and Kimble, 1989). Shortly afterwards, Notch-1 was identified in humans, (Ellisen et al., 1991), followed by three more paralogues (Weinmaster et al., 1992, Uyttendaele et al., 1996, Lardelli et al., 1994). Notch was linked to binary cell fate choices, in which developmentally equivalent cells adopt mutually exclusive cell fates. In *C. elegans* for example, two adjacent progenitor cells in the hermaphrodite gonad (anchor cells) develop normally into one anchor cell (AC) and one ventral uterine precursor cell (VU). Loss-of function *lin-12* mutations produce only AC cells, whereas overactive *lin-12* produces only VU cells (Greenwald et al., 1983). Laser ablation studies revealed that the transformation into a neuroblast inhibited its neighbours from entering the same cell fate (DOE and GOODMAN, 1985), suggesting that Notch has an inhibiting effect on neighbouring cells; the neurogenic phenotype observed earlier, apparently results from an absence of inhibition to differentiate

into neural cells. The physical interaction of Notch with Delta ligand was shown in cell aggregation assays using cells expressing either Notch or Delta (Fehon et al., 1990). The molecular dissection of Notch has since been performed by a myriad of studies which show that Notch is an highly conserved metazoan pathway, with crucial roles in almost all developing tissues, for review see Guruharsha (2012).

1.1.3 Notch - three modes of action

Three different modes of actions (Figure 1.3) were proposed in an attempt to categorise the diverse effects of Notch (Bray, 1998). Lateral inhibition (Figure 1.3A) explains how one out of many equivalent cells can adopt a different cell fate assuming that Notch has a cell fate inhibitory effect on neighbouring cells. In a schematic stretch of cells, the surrounded cell receives more inhibition than cells on the outside of the stretch. This difference is thought to be amplified over time (arrow) and results in the differentiation of cells with less inhibition (grey cells), but not of the strongly inhibited one (white cell) (Greenwald, 1998). In asymmetric cell division (Figure 1.3B), Notch is active during cell proliferation in one cell, but not in the other; this results in the adoption of two different cell lineages.

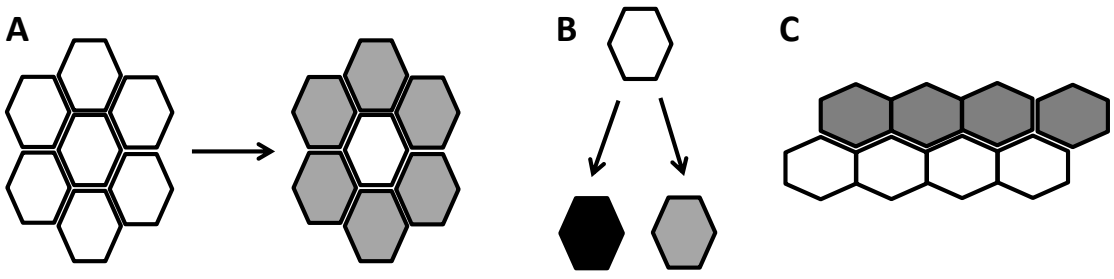


Figure 1.3 Three modes of Notch action. (A) In lateral inhibition, cells from a group of equivalent precursors inhibit their neighbours' ability to adopt a certain cell fate. This inhibitory effect is amplified over time (arrow) and leads to differentiation (grey) or not (white). (B) Asymmetric cell divisions result in cells with different specifications, due to the fact that Notch is activated in one cell (black), but not in the other one (grey). (C) Inductive signalling leads to boundary formation which ensures a different cell fate along a boundary, such as for the dorsal-ventral boundary (grey dorsal cells, white ventral cells) in the *Drosophila* wing. Cells are represented by hexagons, and the distinct shadings indicate different cell fates.

An example is the bristle precursor, in which two subsequent lineage decisions lead to shaft, socket, glia and neuron cells. Only neurons are formed in the absence of Notch (Bray, 1998). Inductive signalling (Figure 1C) explains the ability of Notch to form boundaries, for which a good example is the dorsal-ventral boundary in the *Drosophila* wing.

1.2 The structural basis for Notch signalling

1.2.1 The core pathway and key players

The DNA of Notch was first sequenced in *Drosophila*, revealing that Notch is a trans-membrane protein, consisting of 36 EGF-like domains (Figure 1.4) (Wharton et al., 1985). Since then, the structure and action of many key players involved has been investigated, for review see Gordon (2008) and Chillakuri (2012). In humans, there are four, structurally similar Notch receptors, with 29-36 EGF domains, whereas there are 2 receptors in *C. elegans* with only 11 and 14 EGF domains (called LIN-12 and GLP-1) and 1 receptor in *Drosophila* with 36 EGF domains (Fleming, 1998). There are two main ligand families Jagged/Serrate and Delta, both of which contain a DSL domain. There is one of each, Serrate and Delta, in *Drosophila*, but two Jagged ligands as well as 3 Delta-like ligands in humans.

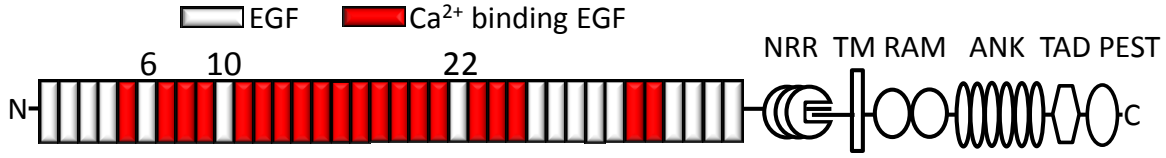


Figure 1.4 Modular organisation of the human Notch-1 receptor. Ca^{2+} binding EGF domains are shown in red, non-Ca^{2+} binding domains in white. The N and C termini, EGF domains 6, 10 and 22, the negative regulatory region (NRR), the transmembrane domain (TM), the RAM domains, the ankyrin-repeats (ANK), the transactivation domain (TAD) and the PEST domain are indicated.

In *C. elegans*, there are three ligands (APX-1, LAG-2, DSL-1). All ligands, with the exception of the *C. elegans* ligands, contain a conserved ~100 residue domain, called MNNL (module at the N-

terminus of Notch ligands). This domain is followed by a DSL domain and a variable number of EGF domains until the transmembrane region. Serrate and Jagged ligands contain a cysteine-rich domain that shows similarity to the von Villebrand factor C domain.

Notch is synthesised as a ~300 kDa precursor protein that is usually cleaved in the trans-Golgi compartment by a furin-like convertase (S1 cleavage). However, it is not clear whether *Drosophila* Notch undergoes this cleavage. After cleavage in the heterodimerization domain (HD), the two subunits of the HD domain assemble together by a non-covalent, but rigid interface, formed by an intertwined α-β sandwich. The extracellular fragment of Notch is subject to N- and O-linked glycosylation processes, required for its proper fold and activity (Irvine, 2008, Takeuchi and Haltiwanger, 2010).

Current models indicate that the recognition of DSL ligands by the Notch extracellular domain causes two subsequent cleavage events in the negative regulatory region (NRR) of Notch. As a result, the receptor-ligand complex is internalised by the signal-sending cell, and the Notch intracellular domain (NICD) translocates into the nucleus where it interacts with the CSL and MAM family of transcription factors to replace the repressor complex CBP/p300. This is thought to activate the expression of target genes such as *hes-1.*

1.2.2 Structure of the ligand binding region

As shown in Figure 1.1, ligand binding between the EGF-rich region in Notch and a DSL-ligand is considered to be the first step during receptor activation; this process is called *trans*-activation (de Celis and Bray, 1997). However, the structural basis for this interaction and how it leads to Notch signalling is not understood, for review see Chillakuri (2012); this is partly due to the large size of both receptor and ligand, their disulphide-rich character and complicated glycosylation profiles. Nevertheless, a variety of elegant studies have provided some insight into receptor/ligand recognition, as reviewed here.

Deletion mapping in combination with a *Drosophila* cell aggregation assay indicated that EGF domains 11-12 were sufficient and required for ligand binding, whereas other EGF domains were dispensable (Rebay et al., 1993, Rebay et al., 1991). However, when this assay was extended to signalling studies, EGF11-12 domains were not able to initiate Notch signalling, suggesting that other events are needed to activate Notch (Rebay et al., 1993). Irvine and colleagues reassessed the outcome of this assay, constructing a series of ectodomain fusion fragments of Notch and alkaline phosphatase, the binding of which to S2 cells expressing Delta/Serrate was analysed (Xu et al., 2005). Their results confirmed the importance of EGF11-12 for ligand binding, but indicated that the other EGF domains contributed at least to some extent to the strength of ligand binding. The involvement of EGF10-12 in ligand binding was also verified in a *Drosophila* system (Becam et al., 2010).

Binding sites within the ligands are also not well understood, but previous data pointed towards the need of the MNNL and DSL domains in addition to EGF1 and EGF2 to provide optimal binding (Parks et al., 2006, Shimizu et al., 1999, Cordle et al., 2008a). Figure 1.5 shows the structure of the proposed ligand-binding region in human Notch-1 (EGF11-13) and the DSL-EGF3 region of the Jagged-1 ligand with residues involved in binding indicated (Cordle et al., 2008a). Both form elongated, gently curved structures, with the inter-domain junctions in Notch-1 being stabilised by Ca^{2+} binding as reported before previously for proteins containing tandem repeats of Ca^{2+} binding EGF domains (Hambleton et al., 2004, Werner et al., 2000).

Binding assays were performed using the minimal binding region in human Notch-1 (EGF11-14) to a Delta-like fragment consisting of the N-terminal domain, the DSL domain and the first three EGF-like domains (Cordle et al., 2008b). Surface plasmon resonance analysis resulted in a K_d of 150 µM for this interaction, which was Ca^{2+} dependent. In a subsequent study, the authors also performed mutagenesis studies on full length Notch and ligand, dissecting ligand-binding and activation events. F249A and R253A mutants of the DSL region in *Drosophila*

Serrate, corresponding to residues F199 and R203 in human Jagged-1, had no effect in cell aggregation assays, but did reduce transactivation as tested by ectopic expression in the imaginal wing disc of *Drosophila* (Cordle et al., 2008a). Mutants of other putative Notch binding residues such as those corresponding to F207 (Figure 1.5), showed both defective binding and signalling.

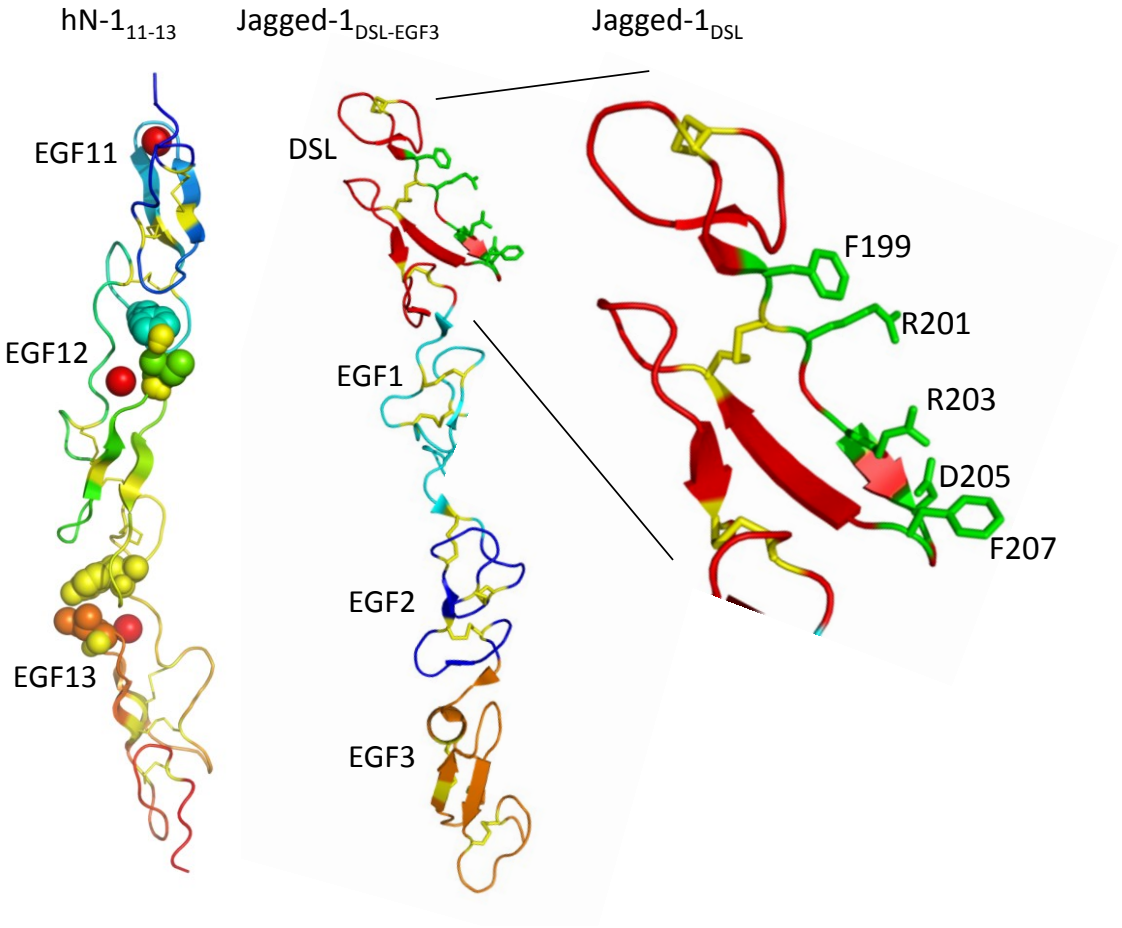


Figure 1.5 X-ray structures obtained for human Notch-1 EGF11-13 (left, PDB ID: 2VJ3) and Jagged-1 DSL-EGF3 (middle and right, PDB ID: 2VJ2). hN-1_{11-13} (left) shows a rod-like, gently curved structure, with interdomain interfaces stabilised by Ca^{2+} (red spheres) and packing of a conserved aromatic residue (spheres, section 1.4.4). Jagged-1 DSL-EGF3 (middle) is linear also and the DSL domain has a novel fold. Residues that were suggested to be involved in ligand binding to the Notch receptor are highlighted as green sticks on the right. Disulphide bonds are shown in yellow sticks.

In-vitro assays with isolated mouse Notch-2 fragment comprising EGF1-15 and Jagged2 ligand fragments indicated a K_d of ~0.7 nM for this interaction in ELISA assays (Shimizu et al., 1999). This suggested that the minimal binding region EGF11-14 was not sufficient to interact

tightly with the ligand. However, the interpretation of this result was not clear since these fragments were Fc-tagged and hence presented as dimers – dimerisation is expected to contribute to substantial avidity effects, for review see Chillakuri (2012). Similarly, the Baker group estimated a K_d of 1.87 nM for *Drosophlia* EGF-like repeats 11-20 and Delta in solid-phase binding assays (Pei and Baker, 2008). Interestingly, a separate Notch fragment, including EGF-like repeats 22-27, competed with the ligand for binding to the EGF11-20 region of Notch. This fragment, included a region referred to as *Abruptex* (EGF24-29) where mutations were shown to lead to enhanced Notch signalling in fly embryos.

Even though the extracellular, EGF-like rich region of Notch might be considered too small to study by single particle electron microscopy (~186 kDa), a recent study challenged this fact and presented a compact structure for Notch (Kelly et al., 2010) (Figure 1.6).

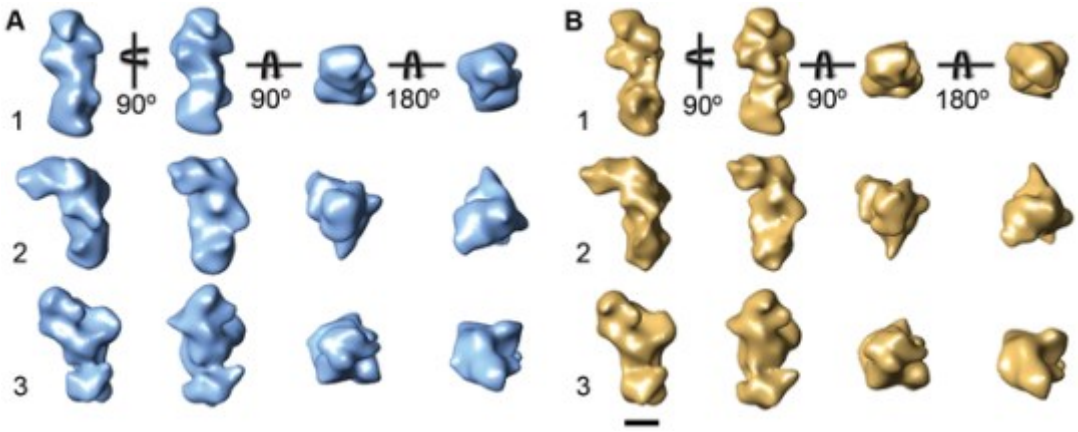


Figure 1.6 Cryo electron microscopy study of the extracellular region of human Notch-1 (A) and *Drosophila* Notch (B). In this study, three distinct conformations were found, identical for both constructs. The error bar represents 5 nm. The image was reproduced from Kelly (2010).

Three different conformers were found, which were at least identical for both human Notch-1 (Figure 1.6A) and *Drosophila* Notch (Figure 1.6B). The length of an EGF domain is around ~3 nm, and a length of 100 nm would have been expected for an extended, rod-like receptor. The compact structures presented, however, are around only 15 nm and dimeric. Since no

characterisation of material used was shown and no evidence of calcium binding presented, as required to validate the native structure of Notch, these data must be viewed as preliminary.

In conclusion, cell aggregation assays point towards EGF11-12 as the core binding site (Rebay et al., 1991), for which structural data indicate a rod-like conformation (Hambleton et al., 2004). The organisation of the other EGF domains within Notch remains unsolved.

1.2.3 *Cis*-interactions

Previous experimental data indicated that Notch can further interact in *cis* with ligands expressed on the same cell-surface (Fehon et al., 1990). This interaction in *cis* has a physiological role in limiting Notch activity, e.g. such as the dorsoventral boundary in the *Drosophila* wing (de Celis and Bray, 1997). Recent work has shed light on the molecular mechanism (Cordle et al., 2008a, Glittenberg et al., 2006, Sprinzak et al., 2010). In a quantitative time-lapse microscopy assay, Notch activation levels were extremely sensitive to Delta expression in *cis,* but less sensitive to Delta expression in *trans* (Sprinzak et al., 2010). Whereas Notch activity correlated with the concentration of Delta in *trans*, a sharp threshold was observed at certain concentrations of Delta in *cis,* above which no Notch activity was observed, independent of the concentration of Delta in *trans*. This is thought to show an ultrasensitive switch between mutually exclusive (sending versus receiving) signaling states that is not dependent on transcription, and can amplify small differences in neighbouing cells without transcription based feedback, as observed during lateral inhibition.

The fact that Notch can interact with Delta either in *cis* or *trans* adds a further layer of complexity in understanding the Notch-ligand binding complex. Are the sites of interaction between Notch and ligand the same during both *cis* and *trans* interaction, or do two completely different interfaces form? Certain residues of the DSL domain were shown previously to be important for both, *trans*-activation and *cis*-inhibition (Cordle et al., 2008a). The Abruptex region

was also implicated in *cis*-interactions (de Celis and Bray, 2000). Nevertheless, the molecular basis of *cis*-inhibition and *trans*-activation complexes is unsolved as is the overall structure of both receptor and ligands.

1.2.4 Receptor activation

Whereas the EGF-rich region of Notch appears key to ligand-binding in canonical Notch activation, the negative regulatory region (NRR) has been shown to be crucial during the next step in receptor activation. Upon ligand binding and in parallel with ligand endocytosis, the NRR region is thought to undergo a conformational change, making the S2 site available for cleavage by the metalloproteinase TACE (TNF-α Converting Enzyme), also known as ADAM17. After the cleavage, most of the extracellular domain of Notch, together with the ligand, is internalised and degraded in the ligand expressing cell. As a consequence, the membrane tethered remainder of the Notch receptor is further cleaved by a γ-secretase (S3 cleavage) and released into the nucleus.

The NRR is situated in between the EGF-rich region and the transmembrane region (Figure 1.4) and consists of three cysteine-rich LIN-12/Notch repeats (LNR) and a heterodimerization domain (HD) which is just next to the membrane and contains the furin (S1) and metalloprotease (S2) cleavage sites (Figure 1.7). The importance of this region was suggested in early experiments. Whereas Notch receptors lacking the EGF-region are non-functional (Rebay et al., 1993, Lieber et al., 1993), the lack of the complete extracellular region leads to constitutive active Notch signalling suggesting that the NRR region is inhibitory. This result was in line with another study on the lin-12 gene of *Caenorhabditis elegans* which showed gain of function effects of mutations within this region (Greenwald and Seydoux, 1990). Usually, most Notch-1 mutations involved in T-cell acute lymphoblastic leukaemia (T-ALL) are located within the NRR region (Weng et al., 2004, Malecki et al., 2006) and lead to constitutive Notch signalling. Figure

1.7 shows the crystal structure of the NRR region of Notch-1. Similar to EGF-like domains, the LNR repeats have little secondary structure and are held together by Ca^{2+} binding and disulphide bonds. Interestingly, Ca^{2+} depletion can lead to receptor activation, independent of ligand-binding events (Rand et al., 2000). The N- and C-terminal HD domains are forming an extensive intertwined interface in an α-β sandwich. The S2 cleavage site within the HD domain is sterically buried by the LNR repeats in a mushroom cap arrangement. Conformational changes after ligand binding are thought to lead to the detachment of the LNR from the S2 cleavage site thus enabling the access of the metalloprotease (S2). It was suggested that the S2 site not only needs to disengage from the LNR repeats A and B, but also detach from the stabilising helix (α3) in the HD domain because of the deep buried active site of ADAM17/TACE metalloproteases (Maskos et al., 1998).

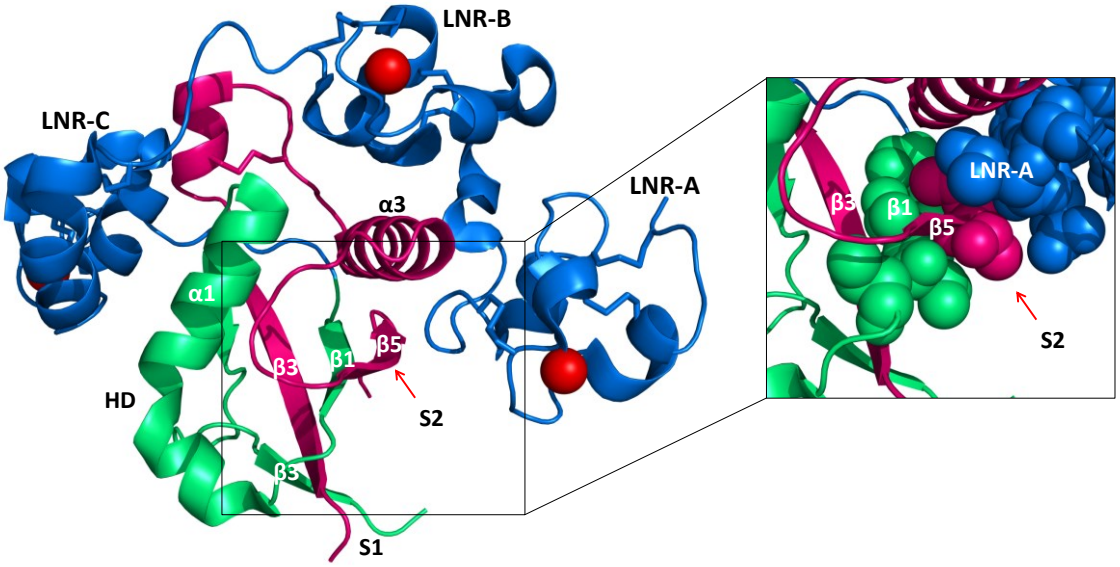


Figure 1.7 The negative regulatory region (NRR) of human Notch-1 (PDB ID: 3I08). The three LNR-repeats (blue) are forming a mushroom-shaped cap, protecting the S2 cleavage site within the heterodimerization domain (HD) (N-terminal green, C-terminal pink) (Gordon et al., 2009a, Gordon et al., 2009b). Note that in this structure the less-defined S1 loop was taken out to enable crystallisation. On the right, the S2 cleavage site is sandwiched by hydrophobic contacts with LNR-A, and the N-terminal HD domain.

This is in line with the unfolding of the HD observed previously (Nichols et al., 2007) and led to the hypothesis of a "mechanotransduction" model, in which ligand-binding to Notch generates a

mechanical force, pulling the LNR-A away from the S2 site; this could explain the requirement of the trans-endocytosis of the Notch-ligand complex into the ligand-expressing cell observed before (Parks et al., 2000) or the fact that soluble ligands are typically unable to activate Notch (Sun and Artavanis-Tsakonas, 1997). Another model involves an allosteric change in the NRR region to occur after the Notch-ligand complex has formed. Interesting results are expected from atomic force microscopy studies (AFM), measuring the force needed to facilitate the liberation of the S2 site (Dey and Szoszkiewicz, 2012, Ahimou et al., 2004). These data might eventually be correlated with the ligand-binding force, calculated on a model of the Notch-ligand interactions.

1.2.5 Transcriptional regulation

The Notch intracellular domain (NICD) is released from the membrane upon S3 cleavage by γ-secretase and consists of 2 RAM domains (recombination binding protein-Jκ-associated molecule), seven ankyrin (ANK) repeats, and a PEST sequence (Gordon et al., 2008). Whereas the 7 ANK repeats and the 2 RAM domains form a complex with the CSL transcription factors (Tamura et al., 1995), the PEST domain regulates the degradation of the NICD; it is thus an important regulator of Notch activity and mutations in this region have been associated with T-cell leukaemia (Weng et al., 2004). The seven ANK repeats, each characterised by a pair of antiparallel helices followed by a β-hairpin that mediates the contact to the next domain (Zweifel et al., 2003) have been shown to be critical for Notch signalling (Rebay et al., 1993, Roehl and Kimble, 1993), but are actually less involved in binding to CSL than the RAM domains (Lubman et al., 2007). The co-activator MAML-1 (Mastermind-like 1 protein), on the other hand, binds to the ANK repeats but not to the RAM domains. Interestingly, MAML-1 binding is cooperative, it binds to the NICD and CSL in a complex, but not to NICD or CSL individually (Nam et al., 2006). The structure of *C. elegans* CSL revealed that three domains, the N-terminal Rel-homology region (RHRN), the central β-trefoil domain (BTD) and the C-terminal

Rel-homology region (RHR-C) are connected by a long β-strand extending from the BTD to the RHR-C domain (Kovall and Hendrickson, 2004) when bound to a 13-mer DNA duplex corresponding to the CSL-binding site in *hes1*. The human CSL complex is shown in Figure 1.8 (Nam et al., 2006)

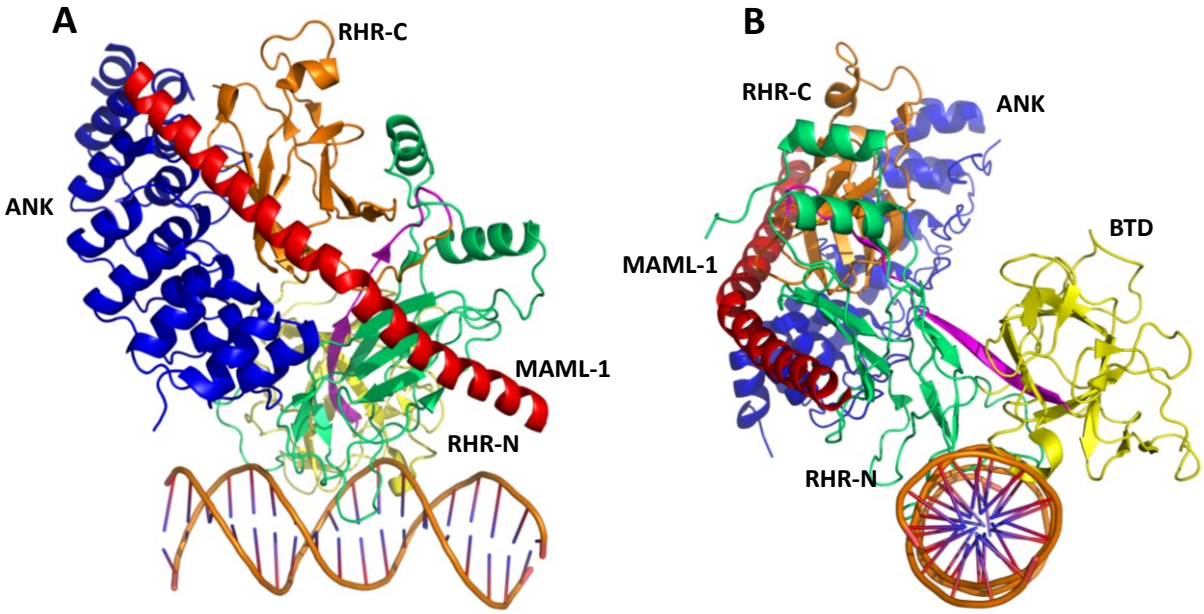


Figure 1.8 (A) Transcription activation complex (PDB ID: 2F8X), (B) side view. MAML-1 in red, ANK repeats in blue, the CSL domains in green. For clarity, the N-terminal region of CSL is green in A and B, the C-terminal in orange and the BTD in yellow. The long β-strand spanning all three domains is indicated in magenta. The complex sits on an 18 base-pair DNA sequence (*hes1* promotor) (PDB cod 2F8X). Note that the RHR-N and the BTD domains are in contact with the DNA.

Interestingly, it has been shown that the ANK repeats can facilitate the formation of a NICD-CSL-MAML1 dimer complex if there are 2 CSL DNA binding sequences present (Arnett et al., 2010).

1.3 Notch in disease

Because of the role of Notch in cell signalling and its involvement in many key biological processes such as apoptosis, proliferation, differentiation, diseases associated with Notch

signalling pathway are highly pleiotropic. Notch dysregulation is not only involved in hereditary diseases, such as Alagille syndrome or CADASIL, but in many types of cancer.

1.3.1 Oncogenic functions

The very first time a human disease was associated with aberrant Notch signalling, was a rare case of T-cell acute lymphoblastic leukaemia (T-ALL) (Ellisen et al., 1991, Reynolds et al., 1987). T-ALL is an aggressive hematopoietic malignancy of thymocytes, mostly diagnosed in children and adolescents and often accompanied by chromosomal translocations of transcription factors in proximity to T cell receptor (TCR) genes. TCR genes are often surrounded by strong enhancers, and the translocation of genes into this region obviously increases their expression. In the case described by Reynolds, the breakage points were the Notch receptor on chromosome 9 and the T-cell receptor β locus on chromosome 7, resulting in constitutive expression of the Notch intracellular domain. Although this specific translocation accounts for only 1% of patients with T-ALL (Ma et al., 1999), a subsequent study showed that more than 50% of T-ALL cases carry Notch activating mutations (Weng et al., 2004). Most of these mutations map into the HD domain of Notch-1 (40-45% of T-ALL patients) and are single amino acid changes that destabilise the interaction between the two subunits leading to ligand-independent cleavage events (Gordon et al., 2007). In addition, several frameshift mutations were found in the PEST domain, suggesting constitutively activated Notch signalling as a mechanism of tumourigenesis. Interestingly, Notch-1 has been shown to be crucial during the maturation of T cell progenitors, their commitment to T cell fate and their assembly in thymocytes and even the T-ALL cells without chromosomal translocations are dependent on Notch signalling (Weng et al., 2004).

Further activating mutations of Notch-1 were found in ~10% of chronic lymphocytic leukaemia (CLL) (Puente et al., 2011, Fabbri et al., 2011) which is the most frequent leukaemia, and 12% Notch-1 mutations were found in mantle cell lymphoma (Kridel et al., 2012). As

opposed to T-ALL mutations, most of the mutations affected the PEST region, but not the NRR region, for review see South (2012).

1.3.2 Tumour suppressor functions

Whereas gain-of function Notch mutations led to the development of T-ALL, a conditional knock-out in mouse skin, as revealed by Radtke and co-workers, increased the rate of skin cancers (spontaneous basal cell carcinoma) significantly, suggesting the role of Notch as a tumour suppressor (Radtke and Raj, 2003, Nicolas et al., 2003). This is consistent with mutations described for chronic myeloid leukaemia (Klinakis et al., 2011). Potential loss-of-function Notch mutations were identified in roughly 10-15% of head and neck squamous cell carcinoma (Stransky et al., 2011, Agrawal et al., 2011) with mutations in Notch-1, Notch-2 and Notch-3. A prevalence of ~75% of loss-of-function Notch-1 and Notch-2 mutations were found in cutaneous squamous cell carcinoma (SCC), one of the most abundant human malignancies, and a smaller number in lung squamous cell carcinoma, but no Notch-1 or Notch-2 mutation was detected in 8 BCC (primary basal cell carcinoma) (Wang et al., 2011). Loss-of-function mutations were mostly nonsense mutations predicted to generate truncated proteins lacking the C-terminal ankyrin repeat domain which is crucial for the activation of target genes, but also missense and in-frame deletions were found in the EGF-rich, N-terminal region. Interestingly, these were not restricted to the minimal ligand binding region EGF11-12, but also to adjacent EGF domains. Often disulphide bonds were affected by the introduction or mutation of a cysteine residue, indicative of potential misfolding. However, also Ca^{2+} binding ligands were affected, such as the D469G mutation (Ca^{2+} binding in EGF12) that was shown to impair ligand binding (Cordle et al., 2008b).

1.3.3 Cancer therapy

Context specific functions for Notch signalling in which Notch action has opposite effects depending on the tissue context (Guruharsha et al., 2012) are also prevalent in cancer – Notch

can act both as an oncogene or tumour suppressor (section 1.3.1 and 1.3.2). Activating mutations were mostly detected in leukaemias and mostly due to mutations in the HD or PEST domain. However, the recent development in whole-genome analysis indicated loss-of-function Notch mutations in many squamous cell carcinomas (SCC) and mutations mostly in the ankyrin domain and to some extent in the ligand binding region. This resembles the structural logic of Notch modulation, in which the ligand binding event as well as the binding event to nuclear transcription effectors are the necessary events leading to Notch activation, whereas the LNR repeats, HD domain and PEST sequence are associated with negative regulatory functions.

The main role of Notch in tumourigenesis is probably unknown, since these tests mainly focused on mutations within Notch itself. Other studies demonstrated that Notch is upregulated in almost all cancers (Groth and Fortini, 2012, Ranganathan et al., 2011). Upregulation of Notch could be a result of glycosylation, proteolysis or other forms of degradation of Notch pathway components. Large scale analysis and systematic approaches to quantify and explain precise relationships are awaited.

Because of early discovery of activating mutations in T-ALL, most therapeutic agents against Notch have so far been targeting oncogenic Notch. Two predominant classes of drugs are currently being tested to reduce the γ-secretase activity. First, γ-secretase inhibitors (GSI) that were initially developed in Alzheimer's therapy, are currently being tested in more than 20 trials against a variety of cancers (Ranganathan et al., 2011). Secondly, antibodies have recently been developed in immunotherapies (Wu et al., 2010, Li et al., 2008, Aste-Amézaga et al., 2010); these have the advantage of being very specific against a potential Notch target and are currently in trials. For example, an antibody developed against the NRR1 region of Notch-1 stabilises the HD-interactions shielding this site from γ-secretase cleavage (Wu et al., 2010). Treatments are also currently being developed to inhibit Notch signalling. For example, Delta-like 4 ligand was upregulated in human tumour vasculature (Patel et al., 2005). Antibodies generated against the

extracellular, receptor binding region in Delta-like 4 are currently being tested in trials (Hoey et al., 2009, Fischer et al., 2011). Altogether, Notch is heavily involved in tumour formation. Whereas most therapies focussed so far on the NRR region, recent studies showed potential in blocking the ligand binding sites between Notch and ligands as in the case of Delta-like 4.

1.3.4 Alagille Syndrome

The Alagille syndrome (AGS) is an autosomal dominant disease, characterised by developmental abnormalities, including liver, heart, skeleton, eye, and face, and affecting around 1 in 50,000 to 1 in 70,000 individuals (Alagille et al., 1975, Danks et al., 1977). This disease was mapped to the Jagged-1 gene (Li et al., 1997, Oda et al., 1997) with mutations spanning across the whole molecule. Some of the mutations caused haploinsufficiency of Jagged-1 and this was the first time that haploinsufficiency of Notch signalling components was associated with human disease, suggesting that dosage effects are not only important in *Drosophila*. Interestingly, mice homozygous for a Jag1 null mutation die during embryogenesis, but heterozygous mice show only eye defects and no further characteristics of Alagille syndrome (Xue et al., 1999). Additional mutations in Notch-2, eventually led to the manifestation of the Alagille syndrome in these mice (McCright et al., 2002). In humans, Notch-2 has also been shown to cause Alagille syndrome (Kamath et al., 2012) with mutations mostly affecting the intracellular region, but also around the putative ligand binding region.

1.4 The cell surface organisation of Notch-1

Whereas the structure of the NRR region (section 1.2.4) and the Notch-intracellular domain (section 1.2.5) have been solved and their function understood, there is still paucity in understanding the structure of the extracellular EGF-like domains and its complex with ligand (Chillakuri et al., 2012). Figure 1.9 shows a schematic picture of hN-1 with available crystal structures. The rod-like, slightly curved shape of EGF11-13, stabilised by interdomain Ca^{2+}

binding (Hambleton et al., 2004, Cordle et al., 2008a), was consistent with other tandem cbEGF domain constructs (Abbott et al., 2007, Smallridge et al., 2003, Downing et al., 1996).

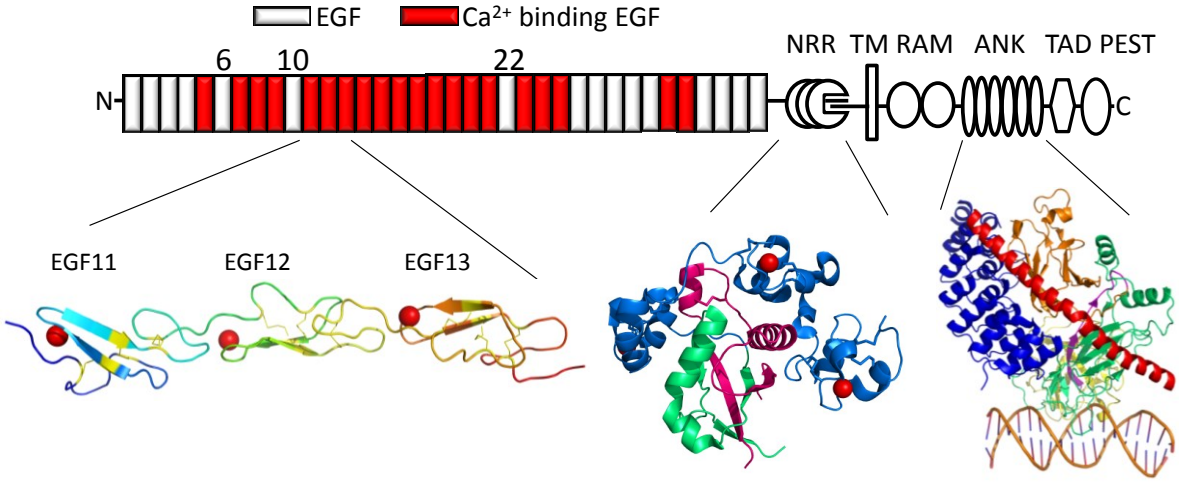


Figure 1.9 Modular organisation of hN-1. The X-ray structures of the NRR region (PDB ID: 3I08) and the ANK region within the transcriptional activation complex (PDB ID: 2F8X) are shown; they reveal a full understanding of their functional role in Notch-signalling. However, so far only the EGF11-13 structure has been solved within the large EGF-like rich region in Notch-1(PDB ID: 2VJ3). It is unclear whether the other EGF domains form a similar organisation or not. Note that Ca^{2+} and Ca^{2+} binding EGF domains are depicted in red, whereas non-calcium binding EGF domains are depicted in white. EGF6, EGF10 and EGF22 were predicted to confer flexibility to the receptor since they do not bind Ca^{2+}.

However, the lack of Ca^{2+} in other domains, as shown for EGF6, EGF10 and EGF22, was predicted to confer flexibility to the receptor (Hambleton et al., 2004). Here, the structure of EGF domains, interdomain contacts and Ca^{2+} binding is reviewed; this is a prerequisite for a further analysis of interfaces formed by non-Ca^{2+} binding EGF domains.

1.4.1 The structure of EGF domains

Epidermal growth factor-like (EGF) domains belong to a family of structurally conserved proteins, abundant in the extracellular space of animals and involved in many cellular processes such as development and cell adhesion. The first solution structure published on the epidermal growth factor itself (Cooke et al., 1987), revealed characteristics of the EGF domain (Figure 1.10). The authors indicated that the structure consisted of 2 domains, shown as the N- (residues 1-32),

and C-terminal (residues 33-48) domain. The N-terminal domain is held together by 4 cysteines in a 1-3 and 2-4 arrangement, whereas two more cysteines form a disulphide bond in the C-terminal domain.

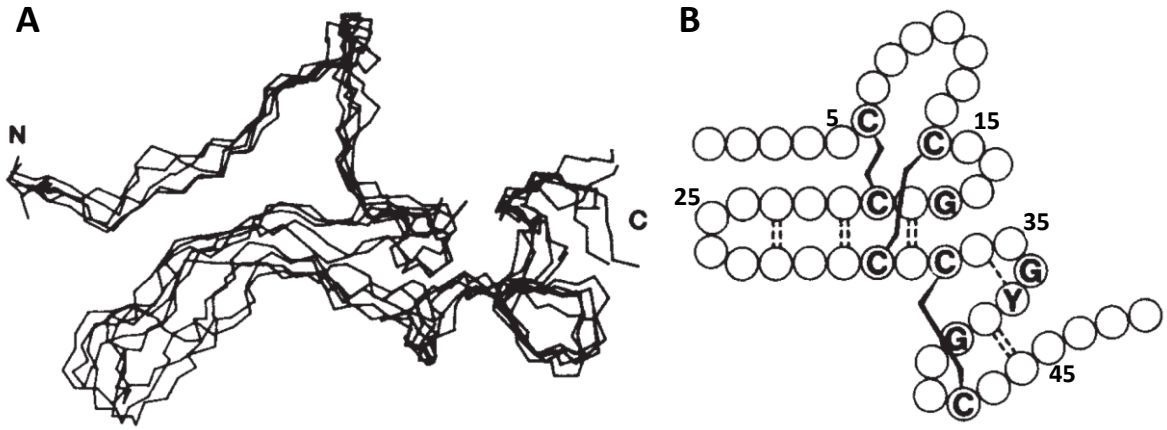


Figure 1.10 Solution structure of the epidermal growth factor (A) and a schematic diagram (B). The N-terminal domain is composed of two disulphide bonds in a 1-3, 2-4 arrangement, spanning from a large β-sheet including a type II β-turn. The C-terminal domain comprises the third disulphide bond (5-6) and a small β-sheet between residues 34-37 and 38-44, and two turns, reproduced from (Cooke et al., 1987).

A large loop in between C6 and C14 showed little NOEs to other regions of the structure, suggesting that this region is less defined. After C14, a kink enables the formation of a long β-sheet, starting from residue 19, with a turn on residues 24-27 and the second part of the sheet with residues 28-32. From here, the C-terminal domain starts with another type II β-turn (residues 34-37) that is followed by a third turn around residue 40 to enable the formation of the third disulphide bond and a minor β-sheet. The authors indicated weak NOEs between residues 5 and 21 and between 3 and 23, respectively, suggesting the possibility of a triple-stranded structure.

In other EGF-like sequences, acidic residues were found at residues corresponding to 1-4 and at residue 22, suggesting that this arrangement could serve to accommodate Ca^{2+}. The authors reported a tight hydrophobic packing, based on an aromatic interface, formed by tyrosines Y13, Y22 and Y29.

1.4.2 Calcium binding EGF (cbEGF) domains

Ca^{2+} binding to EGF domains (cbEGF) was first demonstrated in factors IX and X, protein C and protein S (Persson et al., 1989, Ohlin et al., 1988, Huang et al., 1989, Handford et al., 1990, Dahlbäck et al., 1990). A consensus sequence of D/N-x-D/N-E/Q-x_m-D/N*-x_n-Y/F (where * indicates possible β-hydroxylation, and m/n are variable) is predictive of Ca^{2+} binding (Rees et al., 1988, Handford et al., 1991, Mayhew et al., 1992). The first crystal structure of a cbEGF domain confirmed the EGF fold observed by Cooke (1987) and showed the structural aspects of Ca^{2+} binding (Rao et al., 1995) (Figure 1.11).

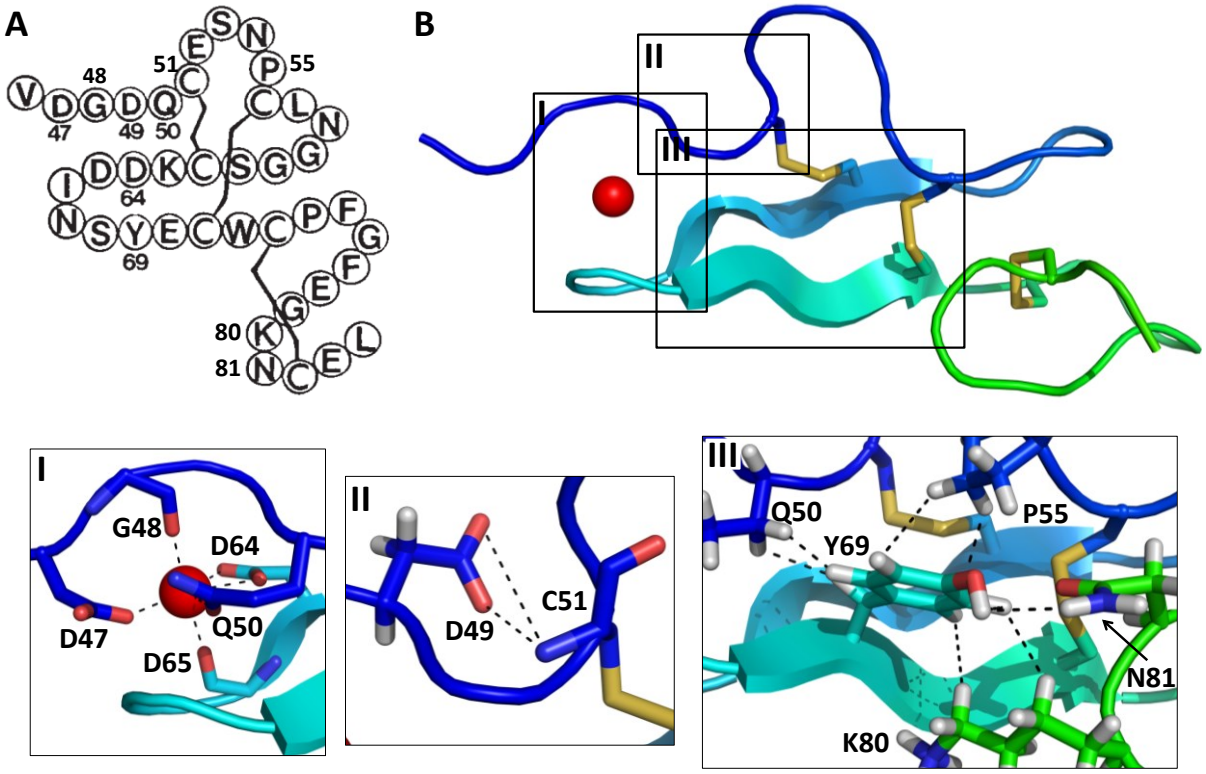


Figure 1.11 X-ray structure of a cbEGF domain in factor IX (PDB ID: 1EDM). Consensus Ca^{2+} binding residues are shown (I). In contrary to what was anticipated, D49 is not directly involved in Ca^{2+} binding, but its carboxylate side chain forms a hydrogen bond with the backbone amide nitrogen of C51, a structural motif known as ASX turn (II). Y69 (III) is not directly involved in Ca^{2+} binding either, but forms a central hydrophobic core with contacts to the side chain of a Ca^{2+} ligand (Q50), the loop region (P55) and the C-terminal domain (N81, K80). Note the close proximity between N81 and P55, a structural aspect not revealed in the schematic diagram (A). Also note that a seventh Ca^{2+} ligand was provided by another EGF domain in the crystal unit (not shown here). Ca^{2+} is depicted in red, disulphide bonds in yellow.

Ca^{2+} ligands (Box I) are arranged in a classic pentagonal bipyramid, with backbone carbonyl oxygen atoms of G48 and D65, and side chain carboxylate groups of D47, Q50 and two side chain carboxylate groups of D64. The seventh ligand was provided by a residue from another EGF domain in the crystal unit, a fact that the authors used to predict a tandem arrangement of cbEGF domains.

The authors indicated that, when compared to other globular proteins, the number of hydrophobic residues is rather low, which is in line with the lack of a strong hydrophobic core. However, some residues belonging to the Ca^{2+} binding consensus sequence showed interesting structural features. The carboxylate side chain of D49, for example, forms a hydrogen bond to the backbone amide of C51 (box II), a structural motif called the ASX turn, which was previously also identified in other Ca^{2+} binding proteins. The last Ca^{2+} binding consensus residue, the aromatic residue (here Y69) two residues before the fourth cysteine, shows a tight packing interaction with the side chains of Q50, P55, K80 and N81, spanning the whole EGF domain. The side chain of Q50 is also a Ca^{2+} binding ligand, and this forms a tight platform for Ca^{2+} binding (Box III).

1.4.3 Posttranslational modifications

Notch is subject to posttranslational modifications by specific enzymes on consensus sequences, with an increasing set of information regarding their biological significance, for review see Takeuchi and Haltiwanger (2010). Like most extracellular proteins, Notch and ligands can have N- and O-linked sugars on the amide-group of asparagine side chains and on the hydroxyl groups of threonine/serine side chains respectively, which are attached during the synthesis in the endoplasmatic reticulum (ER) or in the golgi apparatus. N-linked glycosylation occurs for many proteins on the consensus sequence N-X-S/T, but no specific function for this modification has been observed in Notch signalling (Moloney et al., 2000a, Chen et al., 2001).

Figure 1.12 summarises relevant modifications in Notch. One of the modifications observed from early on in EGF domains is the hydroxylation of the beta carbon of a conserved aspartate or asparagine in the sequence C3-X-D/N-X-X-X-X-Y/F-X-C4 (Stenflo et al., 1988).

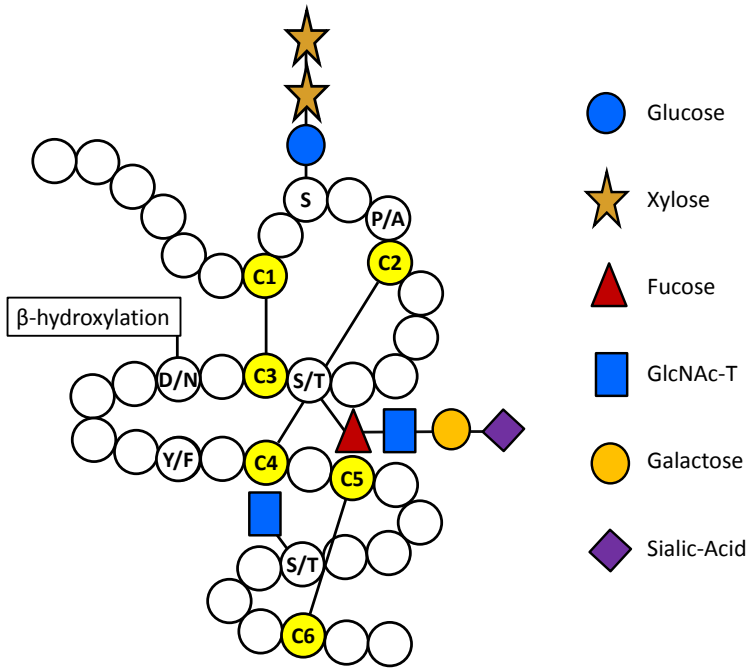


Figure 1.12 Overview of posttranslational modifications in EGF domains of Notch. β-hydroxylation occurs on a D/N residue two residues after C3. Sugar addition is reported for three regions with threonines/serines. Cysteines are highlighted in yellow.

The pattern of hydroxylation is rather difuse, EGF domains are often only partially modified or have the consensus sequence but are not modified (Thim et al., 1988, Stenflo et al., 1988). Since the hydroxylated D/N residue belongs to the Ca^{2+} binding consensus sequence (Handford et al., 1991), a regulatory mechanism was suggested in which beta-hydroxylation promotes Ca^{2+} binding. However, a study performed by Sunnerhagen indicated that there was no change in Ca^{2+} affinity upon hydroxylation (Sunnerhagen et al., 1993), and the fact that the beta carbons point away from the Ca^{2+} ion or the interface suggests an alternative function.

A previous study focused on the physiological role of beta-hydroxylation by knocking-out the responsible enzyme aspartyl β-hydroxylase (BAH). This led to developmental defects similar to mutations causing syndactylism. Interestingly, the BAH null mice had an increased size in

intestinal polyps and number. Up regulation of BAH was also previously reported in different human tumour cell lines (Ince et al., 2000, Lavaissiere et al., 1996). However, no direct link to Notch or its ligands has been demonstrated, indicating that the role of β-hydroxylation on Notch is unknown.

Multiple O-glycosylation sites have been suggested and verified in the Notch receptor (Moloney et al., 2000b). Previously, Rumi was identified as a glycosyltransferase for serines in *Drosophila* Notch with the C1-X-S-X-P-C2 consensus sequence; this modification was shown to be crucial for Notch functioning (Acar et al., 2008). However, the effect was temperature dependent. When growing *Drosophila* Rumi null mutants, loss of Notch signalling and embryonic lethality is observed at 28-30 °C. However, only a mild phenotype was observed at 18 °C and in both cases, no effect on ligand-binding was apparent. Temperature dependent effects had been reported previously in protein folding studies, and this possibly implies a structural effect of O-linked glycosylation. O-linked glucose is further a substrate for xyloltransferases for which two distinct genes, GXYLT1 and GXYLT2, were recently found (Sethi et al., 2010). A human gene, the xyloside xylosyltransferase 1 was also found; this enzyme adds xylol to an alpha-linked xylose of an underlying glucose. A recent study showed that the efficiency of sugar transfer depends on the correct fold of an EGF domain (Takeuchi et al., 2012). Recently, the consensus sequence has been revised in a study of occupancy and efficiency in O-linked glycosylation in mouse (Rana et al., 2011). All of the predicted sites were modified to the trisaccharide Xyl-Xyl-Glc, with differences in tissue specific sites of EGF4 and EGF27 suggesting that some sites are more efficiently modified than others and that this is tissue dependent. Elimination of individual glycosylation sites showed little effect on mouse Notch-1 (mN-1) activation in a cell-based signalling assay, with the exception of EGF 28 (Abruptex region) that did significantly reduce activation.

O-fucosylation has been shown to be crucial for Notch signalling, but the molecular mechanism has not been identified yet. Pofut1 (mammals) and Ofut1 (*Drosophila*) are homologous O-fucosyltransferase enzymes responsible for attaching O-fucose to Ser/Thr in EGF domains with the consensus sequence C2-X-X-X-X-S/T-C3. This enzyme adds O-fucose only to correctly folded EGF domains, for which it was suggested to have a role in quality control (Wang and Spellman, 1998, Luo and Haltiwanger, 2005). Pofut1 knockout mice were embryonic lethal, with a similar phenotype when compared to knockouts of core components of the Notch pathway (presenilins 1 and 2, RBP-Jκ), suggesting it is required for proper Notch functioning (Shi and Stanley, 2003, Donoviel et al., 1999, Herreman et al., 1999, Oka et al., 1995, de la Pompa et al., 1997, Barrantes et al., 1999). Conditional deletion of Pofut1 in mice revealed that Pofut1 is dispensable for early cell fate specifications and for the formation of the three germ layers (Shi et al., 2005), but it was required for the formation of enteric neural crest cells (Okamura and Saga, 2008).

O-fucose is further modified by the enzyme Fringe in the golgi apparatus by the addition of GlcNAc (Moloney et al., 2000a, Brückner et al., 2000). Interestingly, the addition of GlcNAc to O-fucose by Fringe increases Notch binding to Delta but decreases the binding to Serrate (Panin et al., 1997). There is additional complexity in mammals as there are three Fringe enzymes Lunatic, Manic and Radical and it is becoming clear that these play a role in regulating Notch activity, in a tissue-specific context (Takeuchi and Haltiwanger, 2010).

Recently, the solution structure of mouse N-1 EGF12 in the non-Ca^{2+} bound state was obtained (Hiruma-Shimizu et al., 2010). The authors indicated that the addition of GlcNAc-fucose on T464 (Figure 1.13) led to an altered backbone conformation (B). The methyl group of the fucose packs with residues of I475, M477 and P478.

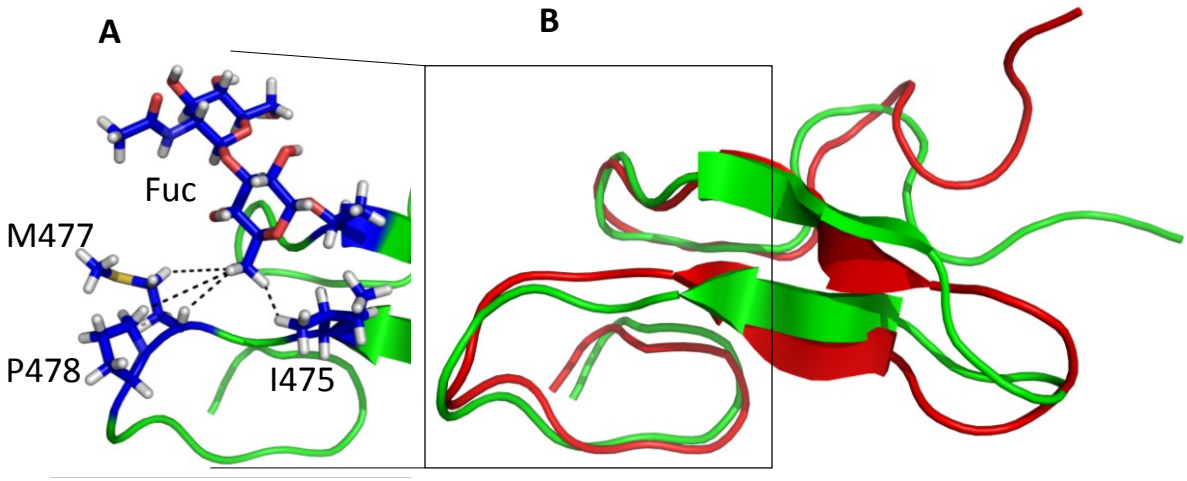


Figure 1.13 shows the structure of mN-1 EGF12. (B) The backbone of EGF12 without any sugar additions (red, PDB ID: 2RR0) is overlayed on the EGF domain with the GlcNAc-fucose (green, PDB ID: 2RR2). (A) The methyl group of the fucose is in a hydrophobic packing core with M477, I475 and P478. This could explain the slight difference in the backbone structures and increased stability. However, it should be noted that these changes were observed in the absence of Ca^{2+}.

1.4.4 cbEGF domain interfaces

Various results showed that the addition of an N-terminal domain to a cbEGF domain increases the Ca^{2+} binding affinity of the latter (Smallridge et al., 1999, Valcarce et al., 1993, Knott et al., 1996). This effect was attributed to hydrophobic packing between a conserved aromatic residue, 4 residues after C5 in the N-terminal domain and hydrophobic residues at the turn of the large β-sheet in the C-terminal cbEGF domain (Downing et al., 1996). Many structural studies on tandem cbEGF domains confirmed this hypothesis and added further details (Saha et al., 2001, Smallridge et al., 2003, Cordle et al., 2008a, Abbott et al., 2007). The interface between EGF11 and EGF12 in hN-$1_{11\text{-}13}$ is shown as an example in Figure 1.14 (Cordle et al., 2008a). Interestingly, G443 (EGF11) provides indirectly a Ca^{2+} binding ligand for EGF12 – its backbone amide has a hydrogen bond to a water molecule which is one of the Ca^{2+} ligands in EGF12 (Box III). This glycine and the indirect Ca^{2+} ligand through a water molecule is identical in the EGF12-13 interface, and a similar arrangement was found in EMR2 (Abbott et al., 2007). The Hα protons of this glycine also pack against the Hα protons of G472 (Box I).

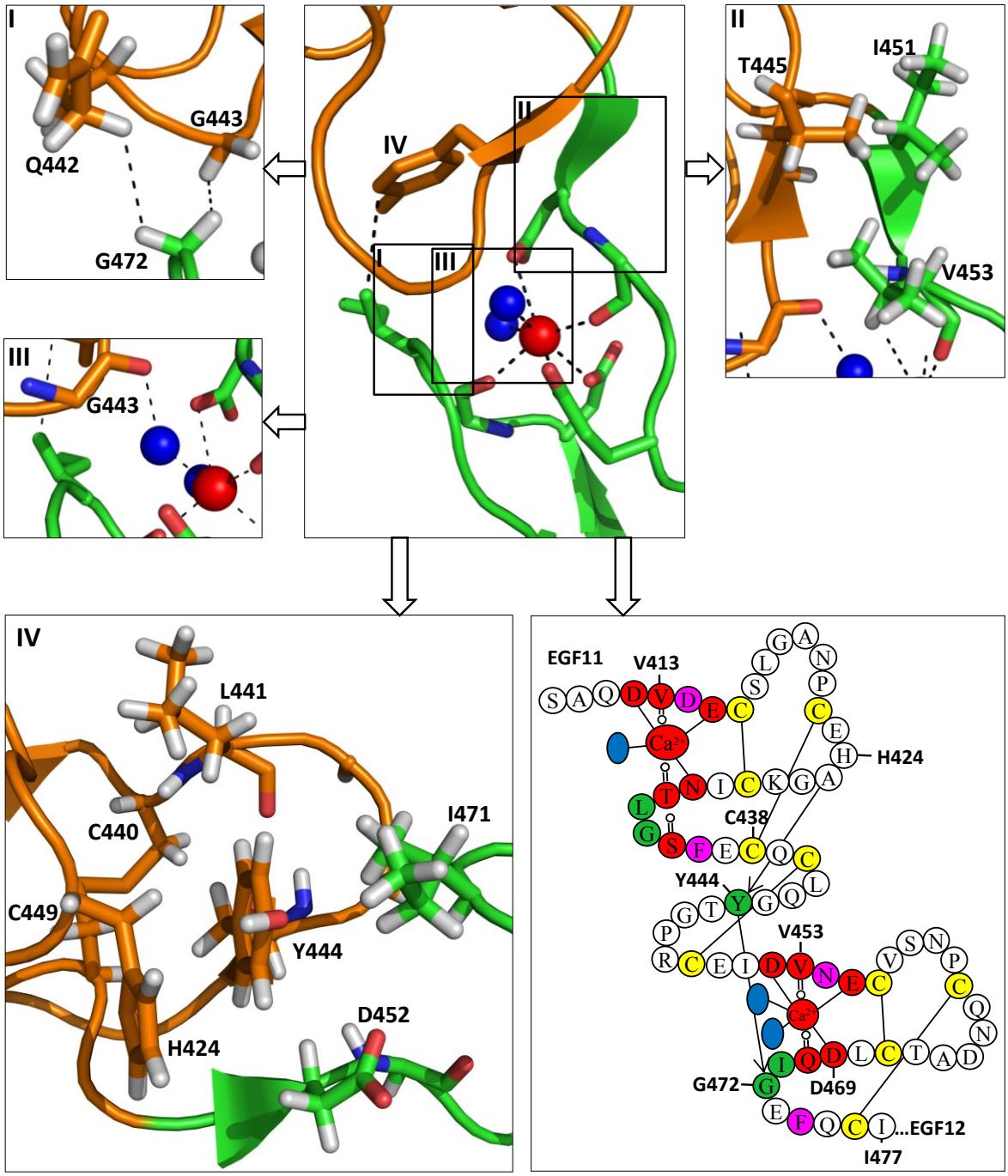


Figure 1.14 Domain interface between Notch-1 EGF11 and EGF12 (PDB: 2VJ3). Several contacts are formed in between both domains. (Box I) Hα protons of G472 (EG12) interact with side chain protons of Q442 and Hα protons of G443. (Box II) The methyl group of T445 (EGF11) is flanked by methyl groups of I451 and V453. (Box III) A water molecule provides one of the Ca^{2+} binding ligands in EGF12 and this water molecule is hold in place by the backbone carbonyl group of G443 (EGF11). (IV) The conserved aromatic packing residue, Y444 (EGF11), packs with many other hydrophobic residues, H424, C440, L441, Y444, C449 (EGF11), as well as with residues of EGF12 (D452, I471). The backbone NH of Y444 has a hydrogen bond to the backbone carbonyl of L441. Hence, Ca^{2+} binding and interdomain packing is rigid and tight. Ca^{2+} and Ca^{2+} ligands are depicted in red, water molecules in blue. In I-IV, residues belonging to EGF11 are coloured in orange, and residues belonging to EGF12 in green. In the schematic diagram on the right hand, cysteines are depicted in yellow, packing residues in green and conserved residues for intradomain stability in purple.

Another contact formed between cbEGFs is demonstrated in box II, where T445 packs against V453 and I451. In box IV, the interactions of the conserved aromatic consensus residue, Y444, are shown. The side chain of Y444 forms hydrophobic interactions with H424 at the N-terminus of EGF11 (schematic diagram) and I471 within the turn of the major beta-sheet in EGF12, the side chains of C440, L441, C449 and D452 are involved in this packing as well. Ca^{2+} binding can be measured by following the NMR chemical shift of aromatic residues in a 1D ^{1}H Ca^{2+} titration (Figure 1.15).

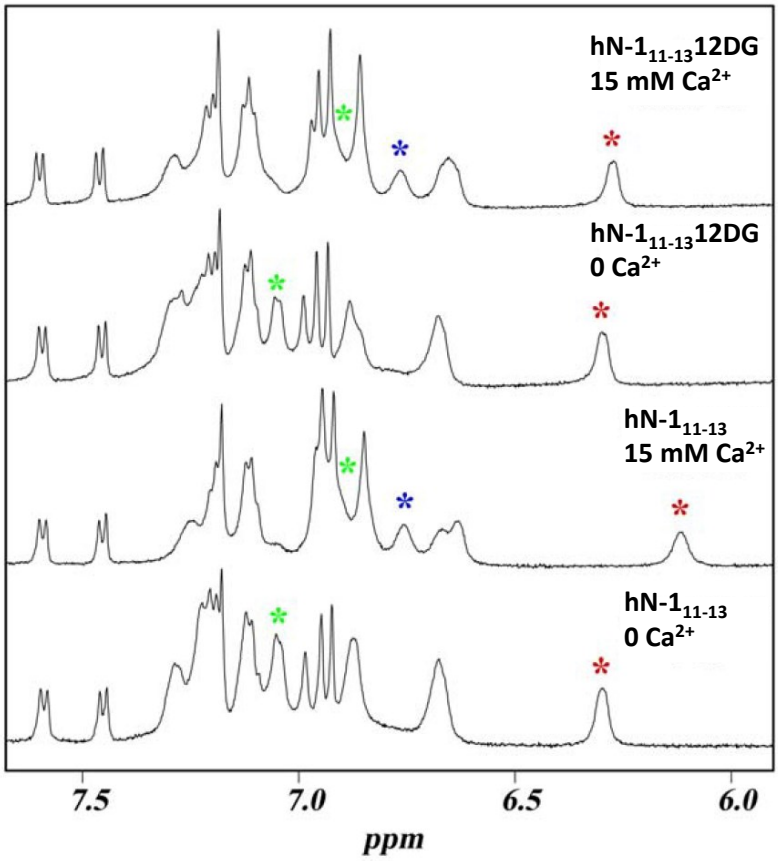


Figure 1.15 1D ^{1}H NMR Ca^{2+} titration of hN-1_{11-13} and hN-1_{11-13}12DG at pH 7.5 and *I* = 0.15. Shown is the aromatic region at 0 Ca^{2+} and at 15 mM Ca^{2+}. Y444, the consensus aromatic packing residue at the C-terminus of EGF11 (red asterisk) shifts with 15 mM Ca^{2+} in the wildtype (hN-1_{11-13}), suggesting Ca^{2+} binding to EGF12 and the formation of a Ca^{2+} dependent interface with EGF11. In the hN-1_{11-13}12DG mutant, however, Y444 does not shift significantly upon addition of 15 mM Ca^{2+}, confirming lack of Ca^{2+} binding to EGF12 and suggesting the absence of an interface with EGF11. Similarly, a shift of F436 (green asterisk) and F512 (blue asterisk) suggests Ca^{2+} binding to EGF11 and EGF13 in both the wildtype and mutant constructs, modified from (Cordle et al., 2008b).

Demonstrated is the effect of knocking-out Ca^{2+} binding in EGF12 by site-directed mutagenesis of an aspartic acid Ca^{2+} ligand D469 to glycine (D469G) (Figure 1.14) (Cordle et al., 2008b). In the wildtype (hN-1_{11-13}), the addition of 15 mM Ca^{2+} leads to the chemical shift of the Hε peak of the consensus aromatic packing residue at the C-terminus of EGF11, Y444 (red asterisk,) suggesting Ca^{2+} binding in EGF12 and the formation of a stable interface with EGF11. In the hN-1_{11-13}12DG mutant, however, no significant shift is observed upon Ca^{2+} addition, confirming the absence of Ca^{2+} binding which suggests a loose interface with EGF11. Ca^{2+} binding in EGF11 and EGF13 is not affected, as shown by the chemical shift of F436 (green asterisk), and F512 (blue asterisk), respectively.

The question arises as to whether the domain orientation observed in the hN-1_{11-13} crystal structure can be extended to the other EGF domains in Notch-1. Table 1.1 shows an alignment of all 36 EGF domains in hN-1; it was assumed that the disulphide bonds are the core folding characteristic of these domains and the cysteines were hence used to have fixed points of overlay. The domain alignment shows the Ca^{2+} binding consensus sequence for most EGF domains. This domain organisation is conserved in *Drosophila*. All EGF domains, whether Ca^{2+} binding or not, possess a conserved aromatic residue packing (Trp, His, Tyr or Phe) in position four after the fifth conserved cysteine, apart from EGF16 and EGF31 where it is a threonine and EGF9 where it is an arginine (section F). The structural consequence of arginine at this position in EGF9 is not known, the packing could as well be maintained along the hydrophobic β/γ/δ protons, or a salt bridge could be possible with the positively charged end. In *Drosophila*, there is a lysine at this position which has similar chemical properties. The conserved aromatic residue packs with the following EGF domain at a hydrophobic residue at the open end of the large β-sheet in the following domain. However, there is considerable variation (Table 1.1, section D) in length and residue type at this open end. This could possibly introduce a slightly different angle between two neighbouring domains. Almost all EGF domains have the conserved glycine, three

residues after C5 (G443), facilitating a β-turn. Exceptions are EGF1, 4, 5, 7, 21, 35 and 36. Of these, only EGF4 and EGF7 were preceded by a Ca^{2+} binding EGF domain. Analysis of the alignment with respect to the packing of T445, I451, V453 (Figure 1.14) showed that these residues are often conserved. Residues corresponding to T445 are mostly threonines.

Table 1.1 Sequence alignment of EGF domains in hN-1. Alignment was performed on conserved cysteines (yellow). Most EGF domains are Ca^{2+} binding. Two additional cysteines were highlighted in red (EGF25 and EGF27); these are not conserved when compared to Drosophila Notch. Boxes A-G indicate regions of sequential characteristics of EGF domains. Not all EGF domains possess the conserved aromatic packing residue (bold, in section F).

EGF	Calcium	A	B	C	D	E	F	G
1	-	RGPR	CSQPGET	CLNGGK	CEAANGTEA	CV	CGGA**F**VGPR	CQD
2	-	PNP	CLSTP	CKNAGT	CHVVDRRGVADYA	CS	CALG**F**SGPL	CLT
3	-	PLDNA	CLTNP	CRNGGT	CDLLTLTEYK	CR	CPPG**W**SGKS	CQQ
4	-	ADP	CASNP	CANGGQ	CLPFEASYI	CH	CPPS**F**HGPT	CRQ
5	**Calcium**	DVNE	CGQKPGL	CRHGGT	CHNEVGSYR	CV	CRAT**H**TGPN	CER
6	-	PYVP	CSPSP	CQNGGT	CRPTGDVTHE	CA	CLPG**F**TGQN	CEE
7	**Calcium**	NIDD	CPGNN	CKNGGA	CVDGVNTYN	CR	CPPE**W**TGQY	CTE
8	**Calcium**	DVDE	CQLMPNA	CQNGGT	CHNTHGGYN	CV	CVNG**W**TGED	CSE
9	**Calcium**	NIDD	CASAA	CFHGAT	CHDRVASFY	CE	CPHG**R**TGLL	CHL
10	-	NDA	CISNP	CNEGSN	CDTNPVNGKAI	CT	CPSG**Y**TGPA	CSQ
11	**Calcium**	DVDE	CSLGANP	CEHAGK	CINTLGSFE	CQ	CLQG**Y**TGPR	CEI
12	**Calcium**	DVNE	CVSNP	CQNDAT	CLDQIGEFQ	CI	CMPG**Y**EGVH	CEV
13	**Calcium**	NTDE	CASSP	CLHNGR	CLDKINEFQ	CE	CPTG**F**TGHL	CQY
14	**Calcium**	DVDE	CASTP	CKNGAK	CLDGPNTYT	CV	CTEG**Y**TGTH	CEV
15	**Calcium**	DIDE	CDPDP	CHYGS	CKDGVATFT	CL	CRPG**Y**TGHH	CET
16	**Calcium**	NINE	CSSQP	CRHGGT	CQDRDNAYL	CF	CLKG**T**TGPN	CEI
17	**Calcium**	NLDD	CASSP	CDSGT	CLDKIDGYE	CA	CEPG**Y**TGSM	CNI
18	**Calcium**	NIDE	CAGNP	CHNGGT	CEDGINGFT	CR	CPEG**Y**HDPT	CLS
19	**Calcium**	EVNE	CNSNP	CVHGA	CRDSLNGYK	CD	CDPG**W**SGTN	CDI
20	**Calcium**	NNNE	CESNP	CVNGGT	CKDMTSGYV	CT	CREG**F**SGPN	CQT
21	**Calcium**	NINE	CASNP	CLNQGT	CIDDVAGYK	CN	CLLP**Y**TGAT	CEV
22	-	VLAP	CAPSP	CRNGGE	CRQSEDYESFS	CV	CPTG**W**QGQT	CEV
23	**Calcium**	DINE	CVLSP	CRHGAS	CQNTHGGYR	CH	CQAG**Y**SGRN	CET
24	**Calcium**	DIDD	CRPNP	CHNGGS	CTDGINTAF	CD	CLPG**F**RGTF	CEE
25	**Calcium**	DINE	CASDP	CRNGAN	CTD**C**VDSYT	CT	CPAG**F**SGIH	CEN
26	-	NTPD	CTESS	CFNGGT	CVDGINSFT	CL	CPPG**F**TGSY	CQH
27	**Calcium**	DVNE	CDSQP	CLHGGT	CQDG**C**GSYR	CT	CPQG**Y**TGPN	CQN
28	-	LVHW	CDSSP	CKNGGK	CWQTHTQYR	CE	CPSG**W**TGLY	CDV
29	-	PSVS	CEVAAQRQGVDVARL	CQHGGL	CVDAGNTHH	CR	CQAG**Y**TGSY	CED
30	-	LVDE	CSPSP	CQNGAT	CTDYLGGYS	CK	CVAG**Y**HGVN	CSE
31	**Calcium**	EIDE	CLSHP	CQNGGT	CLDLPNTYK	CS	CPRG**T**QGVH	CEI
32	**Calcium**	NVDD	CNPPVDPVSRSPK	CFNNGT	CVDQVGGYS	CT	CPPG**F**VGER	CEG
33	-	DVNE	CLSNP	CDARGTQN	CVQRVNDFH	CE	CRAG**H**TGRR	CES
34	-	VING	CKGKP	CKNGGT	CAVASNTARGFI	CK	CPAG**F**EGAT	CEN
35	-	DART	CGSLR	CLNGGT	CISGPRSPT	CL	CLGP**F**TGPE	CQF
36	-	ASSP	CLGGNP	CYNQGT	CEPTSESPFYR	CL	CPAK**F**NGLL	CHI

Sequences of around 250 EGF-cbEGF and cbEGF-cbEGF pairs have previously been analysed and divided into two groups; Class I EGF domains contained one residue in between C6 and the first calcium binding residue of the following domain, and Class II EGF domains two residues (Downing et al., 1996). Whereas the tilt angles were similar, (17° +/- 4°) in class II as shown on the solution structure of hN-1_{11-13} (Hambleton et al., 2004) and 18° +/- 6° in class I EGF domains as shown in the cbEGF32-33 pair of fibrillin-1 (Downing et al., 1996), a large difference in twist angle was found. This was 159° +/- 6° in the cbEGF32-33 structure, but 119° +/- 5° in the hN-1_{11-13} structure. Notch-1 contains mostly EGF Class II domains, with the exception of EGF2, EGF3, EGF4, EGF10 and EGF36. Interestingly, EGF10, adjacent to the ligand-binding EGF11-12 is a class I domain.

In conclusion, most EGF domains are calcium binding and have a sequence similar to the EGF11-13 construct, suggesting a linear, rod-like organisation. The interfaces formed by non-calcium binding EGF domains, however, remain obscure.

1.4.5 Interfaces without Ca^{2+}

Several reports presented interfaces for non Ca^{2+} binding domains. Wolfram Bode and colleagues showed the crystal structure of factor IX which contained a cbEGF-EGF interface (Figure 1.16) (Brandstetter et al., 1995). This construct crystallised in an unusual conformation, with a salt bridge between E78 and R94. F77, the conserved aromatic packing residue had no effect on interdomain contacts at all. It is however possible that this does not reflect the solution states of these domains, since the crystallisation was performed without Ca^{2+}.

Similar, “unusual” interfaces were observed in other proteins of *Plasmodium vivax*, but these contained a different disulphide arrangement (Babon et al., 2007, Saxena et al., 2006). The structure of cbEGF-EGF interfaces or EGF-EGF interfaces therefore remains unknown.

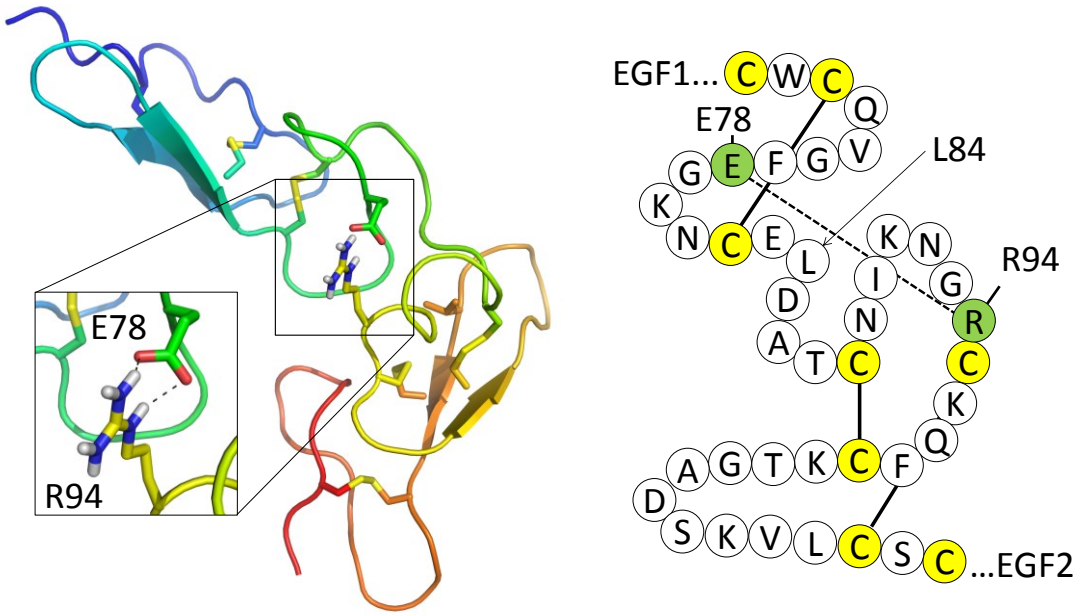


Figure 1.16 Crystal structure of a cbEGF-EGF interface in factor IX (PDB ID: 1PFX). The crystal structure shows a bent in between EGF1 (cbEGF) and EGF2 (EGF) domain and a completely different interface to what was observed for tandem cbEGF interfaces. Highlighted in green are E78 of EGF1 and R94 of EGF2 that form a salt bridge. The first residue of EGF2 is L84. Cysteines are highlighted in yellow.

1.4.6 Summary and aim of thesis

Whereas the downstream consequences of Notch activation have been studied widely, little is known about the early events that lead to the activation of the receptor. Notch and its ligands are transmembrane proteins that consist mainly of epidermal growth factor-like domains (EGF-domains), many of which are Ca^{2+} binding (cbEGF). The crystal and NMR structures of hN-$1_{11–13}$ showed a rigid, rod-like conformation (Cordle et al., 2008a) that is stabilized by interdomain hydrophobic packing and Ca^{2+} binding (Werner et al., 2000). The interfacial interactions between EGF-cbEGF and cbEGF-EGF domain pairs are still to be determined. These interfaces could be extensive like the transforming growth factor β binding protein-like (TB) domain of fibrillin-1 TB4 within the context of cbEGF-22-TB4-cbEGF-23 (Lee et al., 2004), bent with a novel interface like in the factor IX (Brandstetter et al., 1995), or loose with a flexible linker as in TB6-cbEGF-32 (Kettle et al., 1999).

In human Notch-1, three regions next to the ligand binding site contain a non-calcium binding EGF (Figure 1.17 A) flanked by two cbEGFs (EGF 5-7, EGF 9-11, EGF 21-23). These sites could contribute to flexibility within the receptor (Hambleton et al., 2004), a hypothesis tested in this thesis. Therefore, triple or quadruple EGF domain constructs containing EGF6, EGF10 and EGF22 are constructed in a well-established prokaryotic expression system (Cordle et al., 2008a) and refolded *in-vitro*. Available NMR chemical shift assignments of the construct EGF11-13 (Muranyi et al., 2004) are used as a starting point.

A detailed biophysical characterisation of these construct will make it possible to distinguish between a linear (Figure 1.17A) and a non-linear (Figure 1.17B) conformation. The outcome of this research will form the basis for understanding the Notch structure on the cell surface, its interactions with ligands and receptor activation.

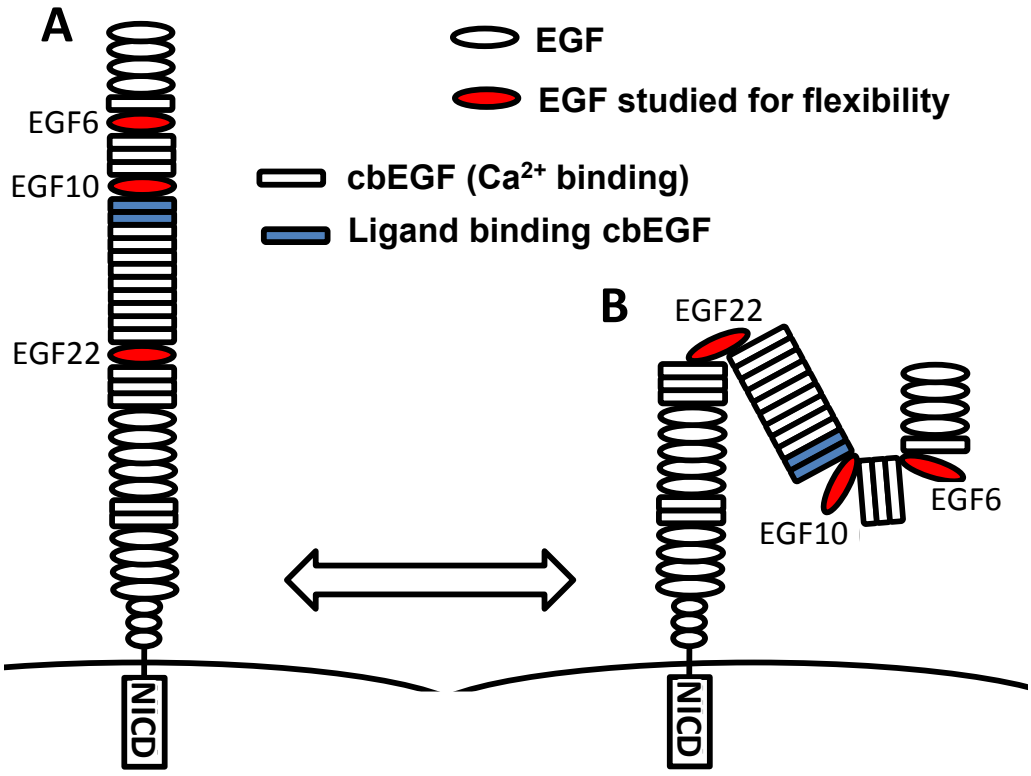


Figure 1.17 Proposed points of flexibility within the Notch receptor. A study of cbEGF-EGF-cbEGF constructs, including EGF6, EGF10 and EGF22, will provide information on whether the Notch receptor adopts a linear (A) or jack-knife (B) conformation. Regions discussed in this report are indicated in the legend, the Notch intracellular domain (NICD) is indicated.

2 Materials and methods

2.1 Amplification and cloning of DNA

Constructs, comprising the regions EGF1-11 (residues 20-451) and EGF20-23 (residues 753-905) of human Notch-1 and EGF4-7 of *Drosophila* Notch (residues 177-329) were amplified from human Tan1 cDNA and *Drosophila* Notch cDNA using *Pfu* polymerase (Stratagene) and primers are shown in Table 2.1.

Table 2.1 Primers used for the amplification of Notch DNA. Forward (F) and backward (R) directions (D) are indicated.

Primer	D	Sequence (5'-3')
EGF1	F	TAG TAG GGA TCC ATA GAA GGA CGA TCA GCA CGA GGC CCG CGA TGC TCC CAG
EGF4	F	TAG TAG GGA TCC ATA GAA GGA CGA TCA GCA CAG GCT GAC CCG TGC GCC
EGF4EKDRO	F	TAG TAG GGA TCC GAT GAC GAT GAC AAA TCA GCA ACG AAG AAT CTG TGC GCA AGT TC
EGF5	F	TAG TAG GGA TCC ATA GAA GGA CGA TCA GCA GAT GTC AAC GAG TGT GGC CAG
EGF7	F	TAG TAG GGA TCC ATA GAA GGA CGA TCA GCA AAT ATC GAC GAT TGT CCA GGA AAC
EGF8EK	F	TAG TAG GGA TCC GAT GAC GAT GAC AAA TCA GCA GAT GTG GAC GAG TGC CAG C
EGF 9	F	TAG TAG GGA TCC ATA GAA GGA CGA TCA GCA AAC ATT GAT GAC TGT GCC AGC
EGF20EK	F	TAG TAG GGA TCC GAT GAC GAT GAC AAA TCA GCA ATC AAC AAC AAC GAG TGT G
EGF21EK	F	TAG TAG GGA TCC GAT GAC GAT GAC AAA TCA GCA ACC AAC ATC AAC GAG TGT GCG TC
EGF4	R	TAG TAG AAG CTT CTA TTA CTG CCG GCA GGT GGG GCC
EGF5	R	TAG TAG AAG CTT CTA TTA CCG CTC GCA GTT GGG GCC AGT G
EGF6	R	TAG TAG AAG CTT CTA TTA TTC CTC ACA GTT CTG GCC GGT G
EGF7	R	TAG TAG AAG CTT CTA TTA CTC GGT ACA GTA CTG ACC TGT CCA C
EGF7 DRO	R	TAG TAG AAG CTT CTA TTA GTC CTG GCA GAA TCT GCC CGT G
EGF9	R	TAG TAG AAG CTT CTA TTA GAG GTG GCA CAGCA GAC CTG
EGF9 BIRA	R	TAG TAG GTC GAC GTG GCA CAG CAG ACC TGT GC
EGF11	R	TAG TAG AAG CTT CTA TTA GAT CTC GCA TCG GGG GCC
EGF23	R	TAG TAG AAG CTT CTA TTA GGT CTC GCA GTT GCG CCC ACT G

BIRA: for constructs with the biotin ligase (BirA) tag at the C-terminus (Cordle et al., 2008a)
EK: for constructs containing the enterokinase cleavage site
DRO: for *Drosophila melanogaster* constructs

The annealing and extension temperature were 55 °C and 72 °C, respectively. Forward primers encoded a *Bam*HI-restriction site, a factor Xa cleavage site and a spacer containing a serine and an alanine. The reverse primers encoded a *Hind*III-restriction site and 2 stop codons. The enterokinase cleavage site was used as an alternative method for removing the His-tag for constructs containing a cryptic factor Xa cleavage site. After *Bam*HI- and *Hind*III-digestion (Roche enzymes), the correct size of the construct was tested on an Agarose gel (2%) in 1x TBE (90 mM

Tris-HCl, 90 mM orthoboric acid, 2.5 mM EDTA) containing traces of ethidium bromide. Constructs of the expected size, as estimated from the 100 bp ladder reference (NEB), were cut out and purified using a gel purification kit (QIAGEN). DNA was ligated into pQE30 vector (Qiagen) with the Rapid DNA ligation kit (Roche) to produce a His-tag fusion protein, and/or containing residues for the BirA-tag. Constructs were transformed into competent NM554 cells using heat shock transformation at 42 °C for 90 seconds. After a recuperation period of 1 h at 37 °C in 2xTY-Broth (Formulation/L: 16g Trypone-B, 10g Yeast Extract-B), cells were plated onto agar plates (100 µg/ml ampicillin). The next day, a single colony was inoculated into 100 ml 2xTY medium (Amp 100 µg/ml). DNA was purified using the QIAGEN Plasmid Purification Kit (Midi). DNA was sent to the DNA sequencing facility in Oxford (Source BioScience LifeSciences).

For the cell aggregation assay, Notch mutants were generated in pBSII (KS+) Notch *Sac*I, a shuttle vector containing *Drosophila* Notch EGF domains 1-22. Mutagenesis was performed using the QuikChange™ Lightning Site-Directed Mutagenesis Kit (Stratagene) and primers shown in Table 2.2.

Table 2.2 Primers used to generate full-length Notch *Drosophila* mutants.

Primer	Direction	Sequence (5'-3')
L504A	Forward	CAG AAC GAG GGA TCT TGC GCG GAT GAT CCG GGA ACG TTC
L504A	Reverse	GAA CGT TCC CGG ATC ATC CGC GCA AGA TCC CTC GTT CTG
Y256A	Forward	G GAT TGC GAC ACC AAG GCA AAG CCC TGC TCG CCA TCG C
Y256A	Reverse	G CGA TGG CGA GCA GGG CTT TGC CTT GGT GTC GCA ATC C
256PPP257	Forward	GAT TGC GAC ACC AAG TAC CCA CCG CCA AAG CCC TGC TCG CCA TCG
256PPP257	Reverse	CGA TGG CGA GCA GGG CTT TGG CGG TGG GTA CTT GGT GTC GCA ATC
256GGG257	Forward	GAT TGC GAC ACC AAG TAC GGT GGC GGT AAG CCC TGC TCG CCA TCG
256GGG257	Reverse	CGA TGG CGA GCA GGG CTT ACC GCC ACC GTA CTT GGT GTC GCA ATC
868GGG869	Forward	GAT TGT GAA AGC AAA ATG GGT GGC GGT GAT CCC TGC GCC AGC AAT C
868GGG869	Reverse	GAT TGC TGG CGC AGG GAT CAC CGC CAC CCA TTT TGC TTT CAC AAT C
BS1	Forward	TCC TGC ACA AGT GTC GGT TG
BS2	Forward	CGT GCA AAT ACG GCG GCA C
BS3	Forward	GCT TGG ACT GCA GCA ACA AC
BS4	Forward	GGC TTT AAA TGC AGT TGT GCC
BS5	Forward	GTG CCC GGT TTT ACG GGT C
BS6	Forward	GAT CCC TGC GCC AGC AAT C

The correct sequence was confirmed using a set of 6 overlapping primers (BS1-6) spanning the region EGF2-23. After digesion with *Bst*EII (EGF2) and *Nhe*I (EGF23), fragments were ligated into the hs-Notch and the pMT-Notch vectors. All constructs were verified as correct by DNA sequencing.

2.2 Expression and purification of protein

Recombinant clones, transformed into *E. coli* BL21, containing a *lac* repressor plasmid (prep4) cells were grown at 37 °C in 4L 2xTY medium (Cordle et al., 2008a), supplemented by 100 µg/ml ampicillin and 25 µg/ml kanamycin. Expression of protein was induced at an optical density of 0.8 (600 nm) by the addition of 2 mM IPTG; cells were harvested after 3 h. For the preparation of ^{15}N labelled protein, the protocol was changed to M9 minimal medium in which ^{15}N ammonium-chloride (CortecNet, Sigma, Cambridge Isotope Laboratories, Inc.) was the sole nitrogen source, as described previously (Muranyi et al., 2004). For triple resonance experiments, additional labelling with ^{13}C uniformly labelled glucose was performed. For the M9 medium, the expression efficiency was optimised to a temperature of 18° C, induction time of ~10 hours, and an IPTG concentration of 100 µM. Cells were spun down at 5000 rpm for 10 minutes and lysed with 6 M guanidine-HCl, 50 mM sodium phosphate, 5 mM β-mercaptoethanol for 1 h at RT. DNA was sheared by sonication for 1 min at 20 Watt (Jencons Ultrasonic Processor) and the cell lysate was spun down at 40,000 rpm for 45 minutes (Beckman L7-55). The supernatant was loaded onto a Ni^{2+} chelating Sepharose column (GE Healthcare) and eluted with buffer containing 50 mM EDTA and 100 mM Tris pH 8.3. The protein was reduced by adding dithiothreitol (DTT) to a final concentration of 0.1 M, 100 mM Tris HCl pH 8.3 for 1h at RT, and then dialysed against 0.1% (v/v) trifluoroacetic acid (TFA) overnight. The solution was filtered through 0.2 µM filters (Millex-GP) and the soluble fraction purified by reverse-phase (rp) HPLC using a Beckman Gold system with reverse phase C8 column. A 20-80% buffer gradient (buffer A: 0.1% TFA in H_2O, buffer B: 80% acetonitrile, 0.1% TFA, 20% H_2O) was applied and protein elution

was monitored at 280 nm. Eluted protein was collected typically at around 40-60% buffer B, and lyophilised.

2.3 *In-vitro* refolding and further purification

HPLC purified protein was oxidised *in vitro* using an oxido-shuffling system (Mayhew et al., 1992). The refold was performed typically at a protein concentration of 0.2 mg/ml, 3 mM cysteine, 0.3 mM cystine, 0.1 M Tris-HCl pH 8.3 and optimized using a range of Ca^{2+} concentrations (0-50 mM). The refolding reaction was stopped by the addition of hydrochloric acid to pH 1-3. The oxidised protein solution was filtered through 0.2 µM filters (Millex-GP) before purification and analysis of the soluble fraction by HPLC. After lyophilisation, a further purification step was performed using FPLC ion-exchange chromatography (Mono Q 5/50 GL on an ÄKTA design system) and a 0-100% buffer B gradient over 20 minutes (buffer A: 50 mM Tris-HCl pH 8.3, buffer B: 50 mM Tris-HCl pH 7.5, 1 M NaCl). Constructs containing the factor Xa cleavage site were incubated overnight with one unit of bovine factor Xa (DENZYME) per mg of protein in 50 mM Tris-HCl pH 7.5, 0.1 M NaCl, 1 mM $CaCl_2$ at 37 °C. Constructs containing the enterokinase cleavage site were incubated with an enzyme:protein ratio by weight of 1:1,000,000 overnight at RT. Cleaved constructs were again purified by FPLC ion-exchange chromatography, followed by a final reverse-phase HPLC purification using conditions as described above. For Ca^{2+} titration assays, protein was treated with 3 mM EDTA prior to HPLC purification to remove bond Ca^{2+}. The correct mass and purity of constructs were confirmed by electrospray mass spectrometry, SDS-PAGE under both reducing and non-reducing conditions and by 1D-NMR spectroscopy.

2.4 Determination of protein concentration

Protein concentrations were determined using a Shimadzu UV mini 1240 spectrophotometer, or alternatively a nano-drop (ND-1000, Thermo Scientific). The extinction

coefficients were calculated with Ext(prot) = 1490x(Tyr) + 5500x(Trp) + 125x(Cys), where Tyr, Trp and Cys are the numbers of tyrosines, tryptophanes and half-cystines in the protein construct.

2.5 SDS-PAGE

Protein samples were analysed by reducing and non-reducing SDS-PAGE. 16% acrylamide gels with a 1.5 mm thickness were prepared according to Laemmli et al., (1970) with a separating gel (16% 29:1 acrylamide:bisacrylamide, 0.1% SDS, 0.38 M Tris-HCl pH 8.8, 2% APS, 0.2% TEMED) and a 6% stacking gel (6% 29:1 acrylamide:bisacrylamide, 0.25 M Tris-HCl pH 6.8, 0.1% SDS, 2% APS, 0.2% TEMED). The separating gel was overlayed with isopropanol to exclude oxygen diffusion and to ensure a flat interface to the stacking gel. The isopropanol was washed off carefully, before the stacking gel was poured. Protein was diluted in Laemmli's protein buffer (Laemmli, 1970) with (reducing) or without (non-reducing) 5% β-mercaptoethanol, heated at 95 °C for 5 min and run on the gel with tris-glycine electrophoresis buffer (0.1% SDS, 30 mM Tris-HCl, 250 mM glycine ph 8.3) at 180 Volt for 60 min. Gels were stained with R250 Coomassie brilliant blue (Sigma) in 10% methanol and 10% acetic acid for 1h at RT. Gels were destained overnight in 10% methanol and 10% acetic acid.

2.6 Ca^{2+} chelator competition assays

Ca^{2+} dissociation constants of high-affinity Ca^{2+} binding sites (<10 µM) were determined by competition with the chromophoric chelator 5,5-Br_2BAPTA as described previously (Linse et al., 1991). In this assay, both protein and chelator were diluted in Ca^{2+}-free buffer (5 mM Tris, pH 7.5, 150 mM NaCl) to a concentration of 20-30 µM. The absorbance at 263 nm was monitored during titrations with a Ca^{2+} stock solution (5 mM Tris, pH 7.5, 150 mM NaCl, 1 mM $CaCl_2$) using a Shimadzu UV mini 1240 spectrophotometer. At this wavelength, the absorbance of 5,5-Br_2BAPTA is dependent on the binding of Ca^{2+}. Background Ca^{2+} levels of the protein/chelator solutions were calculated by extrapolating the linear region at the start of each titration to the

y-axis. The absorption due to the protein was calculated by subtracting the absorption values of chelator only solution and protein solution, assuming that unlike chelator, the absorbance of protein is independent of Ca^{2+} binding. The free Ca^{2+} concentration was calculated by the Newton-Raphson method and dissociation constants were calculated as described previously using in-house software (Linse et al., 1991, Stenberg et al., 1997). To estimate the error in the measured K_d values, each titration was repeated at least three times with independently generated material, and the error represented as the standard deviation (SD) (Suk et al., 2004).

2.7 NMR experiments

Experiments were performed on home-built spectrometers incorporating Oxford Instruments Company magnets with operating frequencies of 500 MHz, 600 MHz, 750 MHz and 950 MHz, and equipped with home-built triple-resonance pulsed-field gradient probeheads. Spectrometers were controlled with the GE/OMEGA software. Alternatively, a 500 MHz Bruker Avance spetrometer with attached cryoprobe was used. The temperature was set to 25 °C and left to equilibrate for at least 10 min. The spectrometer was locked to the D_2O frequency and shims were optimised. The ^1H transmitter frequency was set to the water frequency and this served as a chemical shift standard at 4.75 ppm.

2.8 ^1H NMR Ca^{2+} titration studies

After lyophilization, protein was dissolved in 600 µl matrix buffer (99.9% D_2O, 5 mM Tris, 150 mM NaCl) and adjusted to pH 7.5 and a final protein concentration of 300-600 µM. Experiments were performed in standard NMR tubes (Wilmad) on a magnet with a ^1H operating frequency of 500 MHz (Omega). Ca^{2+} in matrix buffer was added at different concentrations and small volume of 1-2 µl, until all peaks stopped shifting. ^1H 1D and 2D NOESY experiments were collected; 1D spectra were processed using Felix 2.3 (Accelrys, Inc., San Diego, CA).

2.9 ^{1}H-^{15}N and triple labelled NMR experiments

All experiments were performed in Shigemi-tubes (Shigemi Inc.) with a protein solution of 250-300 µl and a protein concentration of 1 mM, adjusted to pH 6.1. Samples contained 5% D_2O, traces of sodium azide to prevent bacterial growth and 30 mM Ca^{2+} with the exception of 80 mM Ca^{2+} for the hN-1_{5-7} construct. 3D experiments were performed typically at a protein concentration of 1 mM, which was high enough to buffer the pH at 6.1. The temperature was set to 25 °C and the water peak was used as a reference at 4.75 ppm. Table 2.3 summarises parameters for experiments used. 2D and 3D spectra were processed with NMR Pipe (DELAGLIO et al., 1995) and analysed using the CcpNmr Analysis software (Vranken et al., 2005).

2.10 RDC experiments

Isotropic RDC datasets were collected with sample conditions as described (section 2.9) and an in-phase and anti-phase J-coupled NMR pulse program (Ottiger et al., 1998). Anisotropic conditions were generated by the addition of a hexanol and alkyl-poly(ethylene glycol) mixture (Ruckert and Otting, 2000). Alkyl-poly(ethylene glycol) are denoted as CmEn with m representing the number of carbons in the n-alkyl group and n, representing the number of glycol units in the poly (ethylene glycol) moiety. A 1 ml stock solution of 15% C12E6 (Dodecylhexaglycol, Sigma) was prepared by gently melting C12E6 in a warm water bath (~60 °C) and transferring 150 µg (~150 µl) into an eppendorf tube containing 835 µl pre-warmed H_2O. At a molar ratio between C12E6 and hexanol of 0.64, the system adopts a lyotropic liquid crystalline phase (L_α); the hydrophobic n-alkyl chains aggregate and their hydrophilic poly(ethylene glycol) headgroups point towards the water phase, generating a lamellar phase. This phase is clear but shows a slight opalescence (Figure 2.1). The exact balance between C12E6 and hexanol is crucial; at a hexanol concentration below the optimum, the lamellar phase is unstable and the solution cloudy, whereas above the optimum droplets of hexanol are forming because of its insolubility.

Table 2.3 NMR experiment parameters. SF is the spectrometer frequency, rd the recycle delay, sw the spectral width, mix.time the mixing time, increments are indicated ascomplex increments.

Experiment	SF (MHz)	Blocksize	rd (s)	Direct ^{1}H		Indirect ^{1}H		^{15}N		^{13}C		mix.time
				scans	sw	increments	sw	increments	sw	increments	sw	
^{1}H (1D)	500.07^{O}	4096	1	512	6024							
^{1}H-^{1}H NOESY (2D)	500.07^{O}	1024	1	128	6024			256	6024			150 ms
^{1}H-^{15}N-HSQC (2D)	949.90^{O}	1024	1.6	32	11905			128	3164.5			
^{15}N edited NOESY-HSQC (3D)	949.90^{O}	512	1.6	8	11905	128	10638.3	32	3164.5			150 ms
^{15}N edited NOESY-HSQC (3D)	500.012^{B}	1024	1	8	6666.7	128	6666.67	32	1724.14			75/150ms
^{15}N edited NOESY-HSQC (3D)	600.151^{O}	512	1.2	8	6514.7	128	6724.95	32	2000			150 ms
^{15}N edited NOESY-TOCSY (3D)	600.151^{O}	512	1.2	8	6514.7	128	6724.95	32	2000			40-80ms
^{15}N edited NOESY-TOCSY (3D)	949.90^{O}	512	1.6	8	11905	128	10638.3	32	3164.5			40-80ms
^{1}H-^{15}N NOE (2D)	749.95^{O}	1024	>3s	>96	9434			128	2500			
IPAP ^{15}N-^{1}H HSQC (2D)	598.48^{O}	1024	1.2s	>96	8000			128	2252.25			
HNCO (3D)	500.07^{O}	512	1	32	6668.9			32	1666.67	64	1736.1	
^{1}H-^{15}N-HSQC (2D)	500.012^{B}	1024	1	16/32	6666.7			128	1724.14			
^{1}H-^{13}CHSQC (2D)	500.012^{B}	1024	1	8	6666.7					200	8802.8	
HN(CA)CO (3D)	500.012^{B}	512	1	16	6666.7			28	1724.14	50	1886.2	
HBHA(CO)NH (3D)	500.012^{B}	512	1	8	6666.7			28	1724.14	89	4761.9	
CBCA(CO)NH (3D)	500.012^{B}	512	1	16	6666.7			28	1724.14	50	7462.7	
HNCA (3D)	500.012^{B}	512	1	16	6666.7			28	1724.14	50	3521.1	
HCCH-TOCSY (3D)	500.012^{B}	512	1	8	6666.7	128	6666.67			43	8802.8	
^{13}C-NOESY-HSQC (3D)	500.012^{B}	512	1	8	6666.7	128	6666.67			52	8802.8	80/150 ms

^{B} experiments performed on Bruker spectrometer

^{O} experiments performed on Omega spectrometer

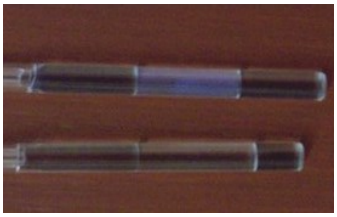


Figure 2.1 Lyotropic liquid crystalline phase in a Shigemi tube. The bottom of a Shigemi tube filled with a standard NMR sample in H_2O is shown on the bottom. On the top, the lyotropic liquid crystalline phase (L_α phase) after the addition of a C12E6/hexanol mix; this phase is still transparent with a slight opalescence.

However, the measurement of small quantities of both C12E6 and hexanol is not straightforward because of the high viscosity of C12E6 and the stickiness of hexanol. Therefore, hexanol was added step by step to the solution until the formation of the lamellar phase was initiated (~60 µl). The minimum amount of C12E6 to form a lamellar phase in the presence of protein solution is around ~2%. After the addition of the C12E6 stock solution to the protein solution, a further addition of around 0.5 µl of hexanol was required. The strength of alignment was measured using residual quadrupolar coupling of the D_2O signal and this was measured before and after the acquisition of the anisotropic dataset.

2.11 Calculation of torsion angles

Torsion angles of protein structures were calculated using XPLOR version 3.8 software (Brünger et al., 1998). The program TALOS+ was used to predict torsion angles from chemical shifts obtained in this study (Shen et al., 2009). Chemical shifts were exported using the CCPNMR software and then uploaded on the webserver.

2.12 Crystallisation

Proteins were purified to homogeneity by HPLC and then freeze dried. Different conditions and preparations of protein solutions were used to set up crystallisation trays. The optimal protein concentration for crystallisation was approximated in tests by adding 1 µl of protein solution to 1 µl of a 20% PEG 4000, 0.2 mM imidazole malate pH 7.0 solution. The formation of

aggregates was observed under a microscope. The protein concentration was adjusted for aggregation to happen between 2-5 minutes. Crystals were grown at 21 °C by the sitting drop vapour diffusion method and drop volumes of 0.2 µl. Trays were set-up using an Oryx nano crystallisation robot in the laboratory of Prof. Susan Lea at the Sir William Dunn School of Pathology, together with Dr Devon Sheppard. Various screens were used, including the Stura Footprint Screen, the MacroSol HT and the JCSG screen (Molecular Dimensions). Initially, proteins were dissolved in a Ca^{2+}-containing solution, as for the crystal structure of hN-1_{4-7} which was solved by molecular replacement. For hN-1_{21-23}, Ca^{2+} was replaced by Ba^{2+}, a heavier atom with similar properties to Ca^{2+}, and the structure was solved by multiple isomorphous replacement (MIR).

2.13 Cell aggregation assay

S2 insect cells were incubated in Schneider's medium (Invitrogen), containing 5% FBS (fetal bovine serum) and 0.5% pen-strep at 25 °C and 5% CO_2. Figure 2.2 summarises the protocol for the cell aggregation assay. At day 0, 6-well plates were supplemented with 1.5 million cells per well and 2 ml of medium. On the following day, a Serrate construct in pUAST containing GFP, was transfected. 12 µl of $CaCl_2$ were combined with 2 µg of DNA and adjusted with sterile water to 100 µl. This solution was added drop by drop under continuous mixing to a solution of 100 µl HBS (HBS of the Calcium phosphate transfection Kit, Sigma). The process of mixing was performed slowly over a period of 2 minutes. The mix was left to incubate for 30 minutes and then added to the 6 well plates. On day 3, Notch was transfected. On day 4, Serrate expression was induced by adding copper sulfate ($CuSO_4$) to the wells with Serrate transfected cells to a final concentration of 500 µM. Notch expression was induced at day 5 by placing the cells for 40 minutes in a 37 °C incubator (heat-shock induction). After 4 hours at RT, Notch and Serrate cells were mixed and put in Eppendorf tubes that were mounted on a rotation device and cells were rotated for 30 min. 500 µl of cell mixtures were pipetted on poly-lysine coated cover slips to

accomplish cell attachments. The coverslips were prepared by washing in a washing step with 1% HCl/70% ethanol for 1 hour at RT. Coverslips were dried and a drop with 0.01% poly-L-lysine was added to the coverslip and dried overnight.

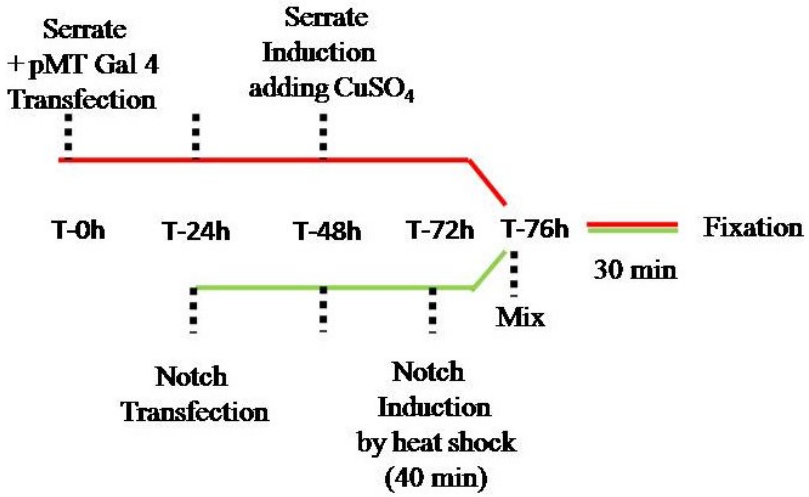


Figure 2.2 Time draft for the cell aggregation assay. The experiment was performed over a period of 4 days (picture adapted from the thesis of Beatriz Hernandez de Madrid, Manchester University).

Cells, not attached were washed away and the attached cells were fixed for 20 minutes with 2% formaldehyde (Thermo Scientific, methanol free) and washed carefully with PBS. Cover slips were blocked overnight in 2ml of 5% skim milk, 0.1% Triton and 0.1% Tween 20 in PBS (blocking reagent). Liquid was removed carefully, and Notch staining was performed with 100 µl of C17.9C6.s antibody (developmental studies hybridoma bank) raised in mice against the intracellular domain of Notch, at a concentration of 1:200 in blocking reagent and an incubation time of 1h at RT. Cover slips were washed carefully with PBS, 5 times, and additional liquid was removed from the wells. Cells were incubated with 100 µl of the secondary antibody, AlexaFluor 594 donkey anti-mouse IG antibody (Invitrogen), diluted 1:200 in blocking reagent for 1h at RT. Cover slips were mounted using 17 µl of vectashield (Vector Labs) on slides and sealed using nail polish. Slides were analysed on a Zeiss fluorescence microscope, using a 40x objective. The number of cells attached to Serrate cells were counted.

3 Characterisation of protein constructs

3.1 Introduction

Whereas the structure of hN-1_{11-13} showed a rigid and rod-like arrangement, stabilised by Ca^{2+}, three central regions of the Notch receptor contain non-calcium binding EGF domains (EGF6, EGF10, EGF22) and hence were predicted to confer flexibility to the receptor (Hambleton et al., 2004). The aim of this thesis is the biophysical and structural characterisation of these sites; the production and initial characterisation of constructs flanking these regions is described in this chapter. EGF domain constructs contain a large number of disulphide bonds, and obtaining milligram quantities of highly pure material is hence challenging. A reliable system for the prokaryotic expression, *in vitro* refolding and purification of disulphide-rich EGF domain constructs has been established in the laboratory of Professor Penny Handford in the last 25 years. Previous examples of similar constructs for which structures have been solved are the human Notch-1 ligand binding region hN-1_{11-13}, the EGF1-2 repeat of the LDL receptor, and the domain pairs cbEGF12-13 and cbEGF32-33 of fibrillin-1 (Hambleton et al., 2004, Cordle et al., 2008b, Downing et al., 1996, Smallridge et al., 2003, Saha et al., 2001).

3.2 Expression of protein

Constructs were cloned using primers described in section 2.1 (Table 2.1). Since the purification protocol was similar for all constructs, the procedure for hN-1_{7-9} has been described in detail with chromatograms for each individual step shown. For all other constructs, SDS PAGE, mass spectrometry and 1D NMR analysis of the final product is shown.

3.3 Refolding and purification of hN-1_{7-9}

The construct hN-1_{7-9} was expressed, purified in reduced form by reverse-phase HPLC and lyophilized as described in Chapter 2. Subsequently, the construct was refolded using an established oxido-shuffling system (Jaenicke and Rudolph, 1989, Mayhew et al., 1992). Trial refolds using a range of Ca^{2+} concentrations (Figure 3.1) showed that in the presence of

10 mM EDTA (A) or in the absence of Ca^{2+} (B), the elution profile was broad suggesting the formation of multiple different oxidised species. The addition of Ca^{2+} resulted in peak sharpening (C-F), suggesting the formation of a smaller number of oxidised species.

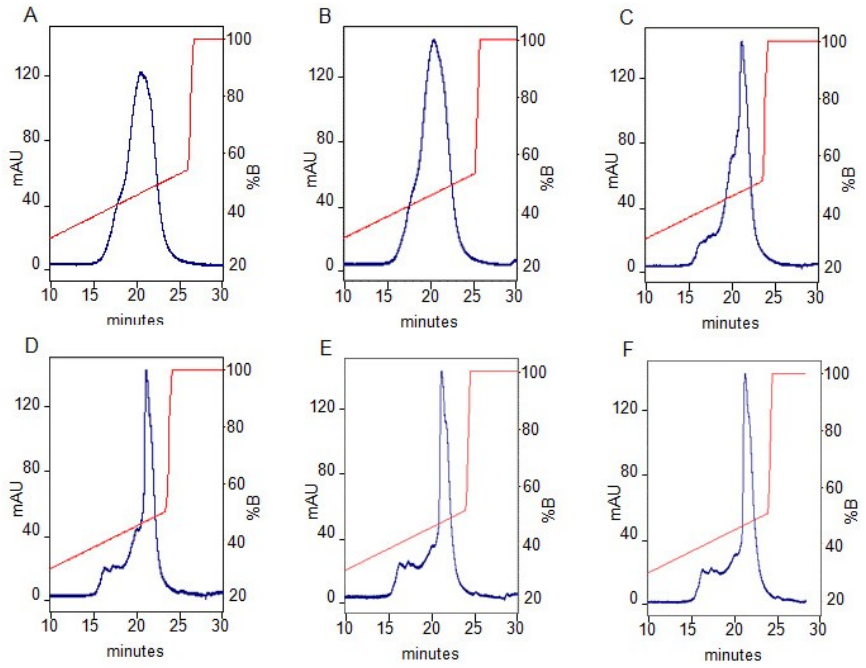


Figure 3.1 HPLC profiles of hN-1_{7-9}, refolded in the presence of 10 mM EDTA (A), without EDTA or Ca^{2+} (B), with 1mM Ca^{2+} (C), 5mM Ca^{2+} (D), 20mM Ca^{2+} (E) and 50mM Ca^{2+} (F). The Ca^{2+} dependence of the refold was confirmed by refolding in the absence of Ca^{2+} (A, B) that gave rise to a broad HPLC profile. Ca^{2+} addition results in peak sharpening (C-F). The red lines show the gradient applied (%B) (see section 2.2).

The pH dependence of the refold was tested as well, but no difference in peak shape was found when using Tris-HCl pH 7.5, pH 8.3 or 9.2 at 5 mM and 30 mM Ca^{2+} (data not shown). The refold condition with 20 mM Ca^{2+} and pH 8.3 (D) was used in subsequent experiments. Similar refold conditions were tested on other constructs, showing that a refold condition of 10-20 mM Ca^{2+} is optimal (not shown).

3.4 FPLC ion-exchange chromatography and further purification

Collected peaks from HPLC samples were dissolved in buffer A (50 mM Tris pH 8.3, 1 mM Ca^{2+}) and applied to a Mono Q column with a buffer B gradient (buffer A, 1 M NaCl) (Figure 3.2A). The main peak usually represented the majority (>70%) of the material. SDS-PAGE analysis

of later eluting material showed that at least some of this was multimeric. The material collected from the main peak was desalted and given a final round of purification by HPLC (Figure 3.2B).

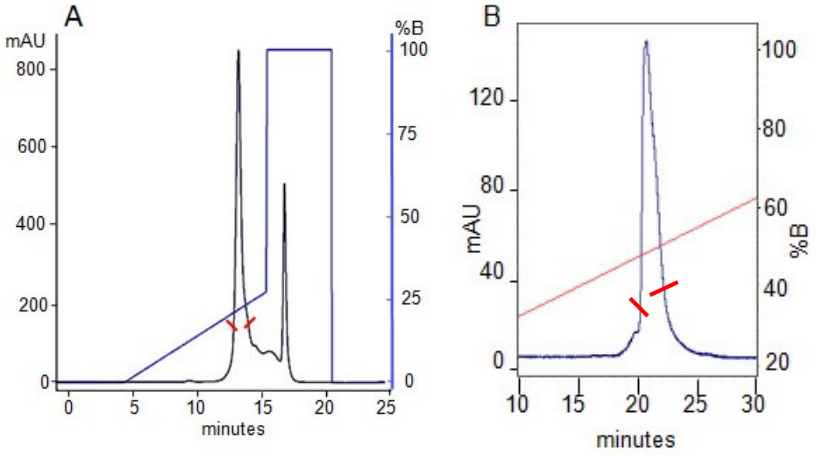


Figure 3.2 Anion-exchange chromatography FPLC elution profile of hN-$1_{7–9}$ (A) and the final HPLC elution profile (B). The blue (left) and purple (right) lines show the gradients applied (%B), short red lines indicate peak collection.

Most oxidised constructs were subjected to His-tag cleavage by factor Xa or enterokinase (as described in section 2.3), followed by a further purification with FPLC and HPLC. However, a cryptic cleavage site between R365 and T366 in EGF-9 susceptible to Factor Xa digestion was identified (un-published). Similarly, EGF7 contained a potential cryptic enterokinase cleavage site. Therefore, the hN-$1_{7–9}$ construct was not cleaved and was subject to further analysis as a His-tag fusion.

3.5 SDS PAGE and Mass spectrometry

All constructs were purified as described in sections 3.4 and 3.5. Most constructs, however, were subject to an enterokinase or factor Xa cleavage as described in section 2.3. After cleavage, an additional purification step with FPLC and HPLC was performed. Table 3.1 shows a summary of constructs generated in this thesis. Constructs hN-1_{11-13} and hN-1_{11-13}12DG were described before and the mass spectrometry analysis not repeated (Cordle et al., 2008a, Cordle et al., 2008b).

Table 3.1 Status of constructs generated in this thesis.

construct	status	cleavage
hN-1_{1-4}	did not refold	-
hN-1_{1-5}	did not refold	-
hN-1_{4-5}	successful	factor Xa
hN-1_{4-6}	successful	factor Xa
hN-1_{4-7}	successful	factor Xa
DN_{4-7}	did not refold well	-
hN-1_{5-7}	successful	factor Xa
hN-1_{7-9}	successful	not cleaved
hN-1_{7-9}birA	successful	not cleaved
hN-1_{9-11}	successful	not cleaved
hN-1_{8-11}	successful	enterokinase
hN-1_{21-23}	successful	enterokinase
hN-1_{20-23}	successful	enterokinase

DN: *Drosophila Melanogaster* Notch construct
birA: a tag added at the C-terminus with a stabilising effect in hN-1_{11-13}

Constructs hN-1_{1-5} and hN-1_{1-4} showed no change in the broad HPLC elution profile after refolding under a set of refolding conditions (pH 6-9, 50% glycerol, Ca^{2+}, EDTA), and were not used in this study. *Drosophila* Notch EGF4-7 did apparently refold, but showed 2 sharp peaks; the construct remains to be characterised further. Other constructs were prepared (Figure 3.3).

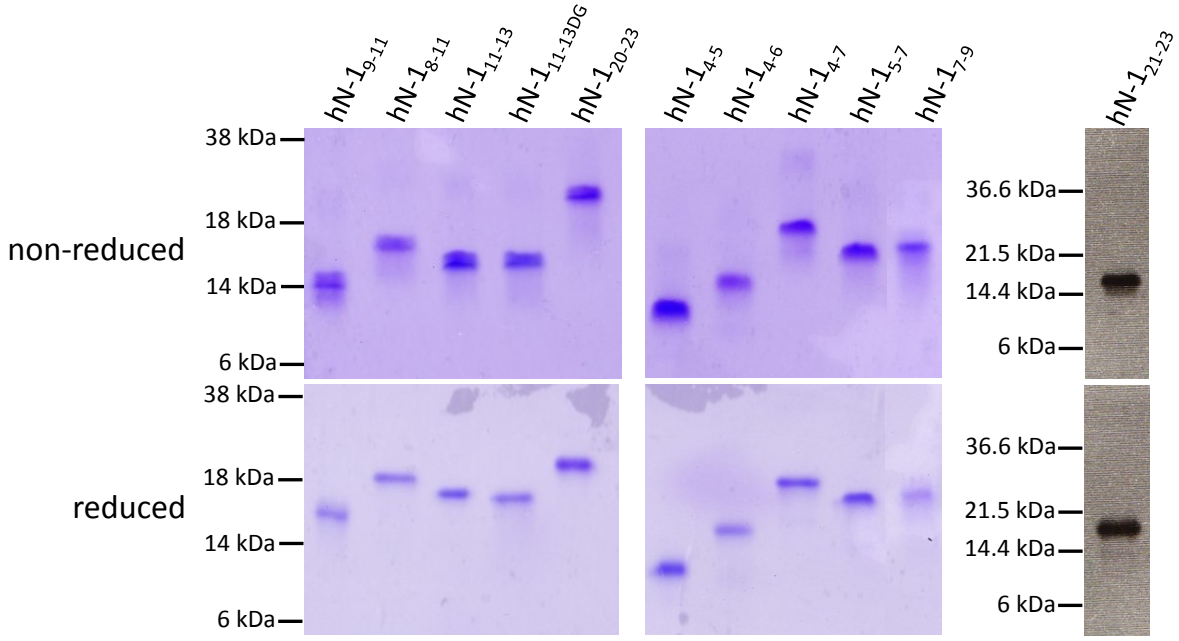


Figure 3.3 SDS-PAGE of different constructs generated in this thesis. Top panels show non-reduced material, whereas the bottom panels present reduced material. All constructs show an acceptable degree of purity under both reducing and non-reducing conditions. The SeeBlue Plus2 Pre-Stained Standard was used as a protein standard, except for hN-1_{21-23} where a Mark12 standard was used.

Figure 3.3 shows the SDS-PAGE analysis of the constructs for which data are presented in this thesis; this was repeated for every protein preparation. All constructs show an acceptable degree of purity (>95%). Table 3.2 shows expected and experimental results from the mass spectrometry (in-house service). The accuracy of the method was tested on the construct hN-1_{11-13}12DG, which is well characterised (Cordle et al., 2008b). All results were within 7 Da of the expected mass.

Table 3.2 Results of mass spectrometry. The largest difference in between expected and experimental mass was for hN-1_{20-23}, with around 6 Da difference. The mass of hN-1_{4-6} was not measured.

construct	Expected (Da)	Experimental (Da)
hN-1_{4-5}	8597.4 (15N)	8596 (15N)
hN-1_{4-6}	12514.9	-
hN-1_{4-7}	16816.3 (15N)	16812 (15N)
hN-1_{5-7}	12618.8	12618
hN-1_{7-9}	14428.7	14432
hN-1_{9-11}	14223.4	14426
hN-1_{8-11}	16608	16612
hN-1_{11-13} 12DG	14768	14769
hN-1_{21-23}	12641	12640
hN-1_{20-23}	16650.4	16656

3.6 1D ^{1}H NMR Analysis

Although a sharp HPLC elution profile as seen in Figure 3.2B and a single band under non-reducing conditions in Figure 3.3 suggest the presence of a single species of refolded material, it is not clear whether this species is correctly folded. Traditionally, Ca^{2+} binding assays have been used to confirm the presence of Ca^{2+} binding sites indicative of a correct arrangement of disulphide bonds in the protein (Handford et al., 1990). However, the presence of non-calcium binding EGF domains requires other evidence. Here, the Ca^{2+} binding properties of various constructs were tested with 1D ^{1}H NMR spectroscopy to see both aspects of secondary structure formation and Ca^{2+} dependence. Figure 3.4 shows the 1D Ca^{2+} titration of hN-1_{7-9}. The presence of downfield shifted Hα at around 5-6 ppm indicated the formation of β-sheet.

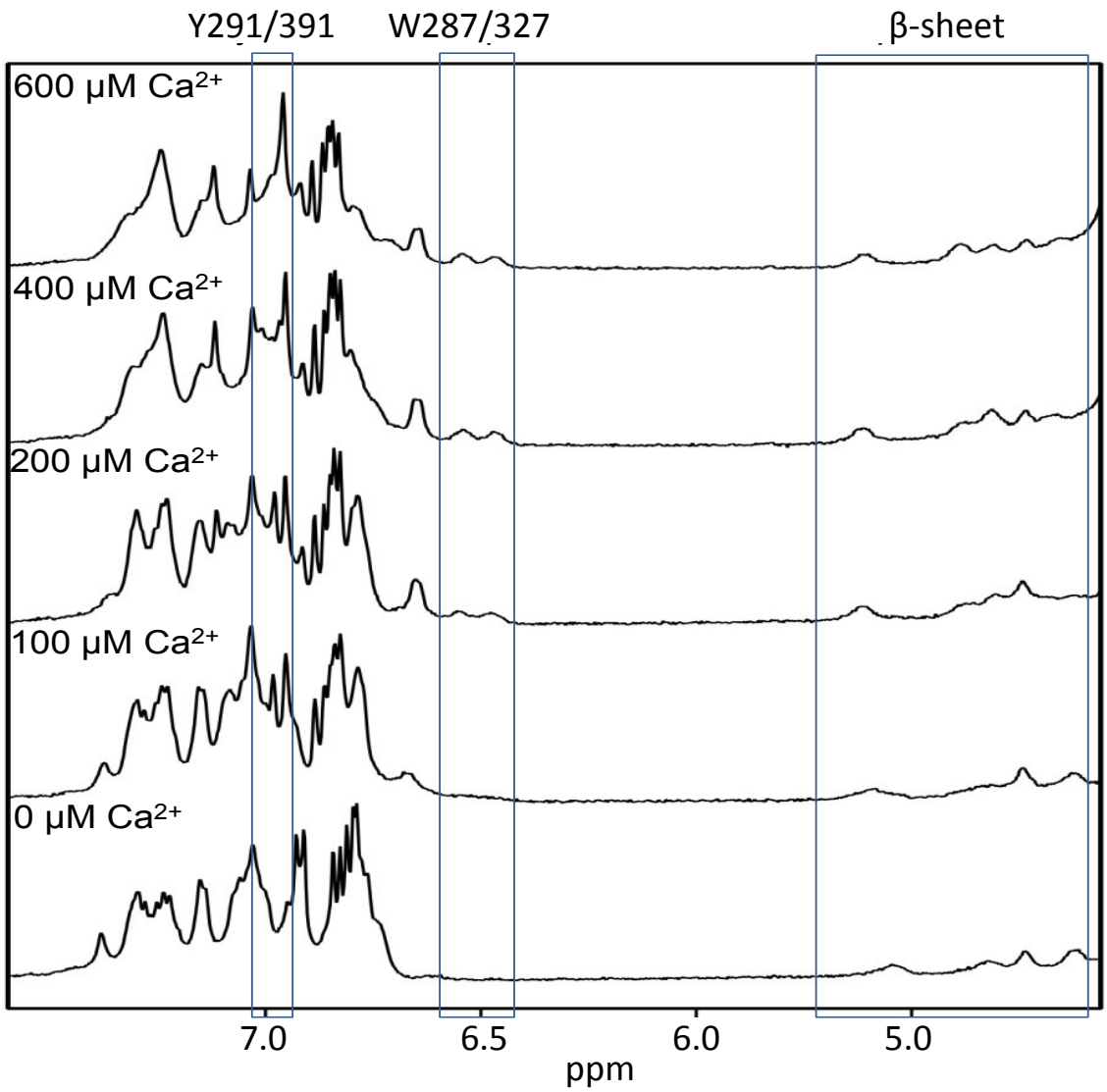


Figure 3.4 1D ^{1}H NMR Ca^{2+} titration of hN-1_{7-9} at pH 7.5 and *I* = 0.15. Note the shift changes upon addition of a low concentration of Ca^{2+}, indicative of high Ca^{2+} affinity and correctly folded protein. The formation of β-sheet, a feature of EGF domains is indicated as well as possible assignments for tryptophans and tyrosines.

Furthermore, large shifts of peaks are observed upon the addition of small amounts of Ca^{2+} (up to 600 µM) indicative of high affinity Ca^{2+} binding sites. The appearance of two peaks around 6.5 ppm, a region usually adopted by tryptophan side chains, suggests the formation of interdomain Ca^{2+} packing (section 1.4.4) either of W287 in EGF7 or W327 in EGF8 (Figure 4.16, section 4.5). A Ca^{2+} dependent tyrosine peak at around 7 ppm belonged either to Y291 or Y319.

In hN-1_{4-7} (Figure 3.5), a tyrosine peak (TyrA) shifts up to a Ca^{2+} concentration of 5 mM indicative of an intermediate affinity Ca^{2+} binding site (schematic diagram in Figure 4.6, section 4.3). Little Ca^{2+} dependence is seen for another tyrosine, the Ca^{2+} binding properties of which is not conclusive from this analysis (TyrB). The construct hN-1_{4-7} contains six histidines and some of them show Ca^{2+} dependence (left hand boxes). A group of peaks at 7.3 ppm is affected by Ca^{2+}; these could possibly belong to a phenylalanine residue (indicated as Phe?). Box and B indicate further peaks affected by Ca^{2+}.

In hN-1_{8-11} (Figure 3.6), Y444, a residue previously assigned in the hN-1_{11-13} construct shows high Ca^{2+} affinity (schematic diagram in section 4.4, Figure 4.11) (Muranyi et al., 2004, Cordle et al., 2008b). A peak at around 6.8 ppm shifts up to 4 mM Ca^{2+}; this peak belongs to Y319, a residue in the N-terminal domain EGF8.

Figure 3.7 shows the spectrum of hN-1_{20-23} for which the crystal structure is currently being refined. Peaks indicated by boxes A-F demonstrate the Ca^{2+} dependence of this construct. Box A indicates a Ca^{2+} dependent histidine, possibly H888, the hydrophobic residue at the open of the large β-sheet in EGF23 that packs against the conserved aromatic packing residue in EGF22 W861 (Figure 3.8). Box F shows a strong peak at around 6.3 ppm. This is possibly Y899 at the minor β-sheet in EGF23, as the peak position matches very well a tyrosine observed in a homologous position in EGF11 (Y444). Box B-E show a mixture of Ca^{2+} dependence at low and high Ca^{2+} concentration, possibly due to Ca^{2+} binding in EGF21 and EGF23, but possibly also due to the Ca^{2+} domain at the N-terminus EGF20 (Y775).

These results suggest that constructs containing cbEGF domains all show Ca^{2+} binding properties, and produce spectra suggestive of folded protein.

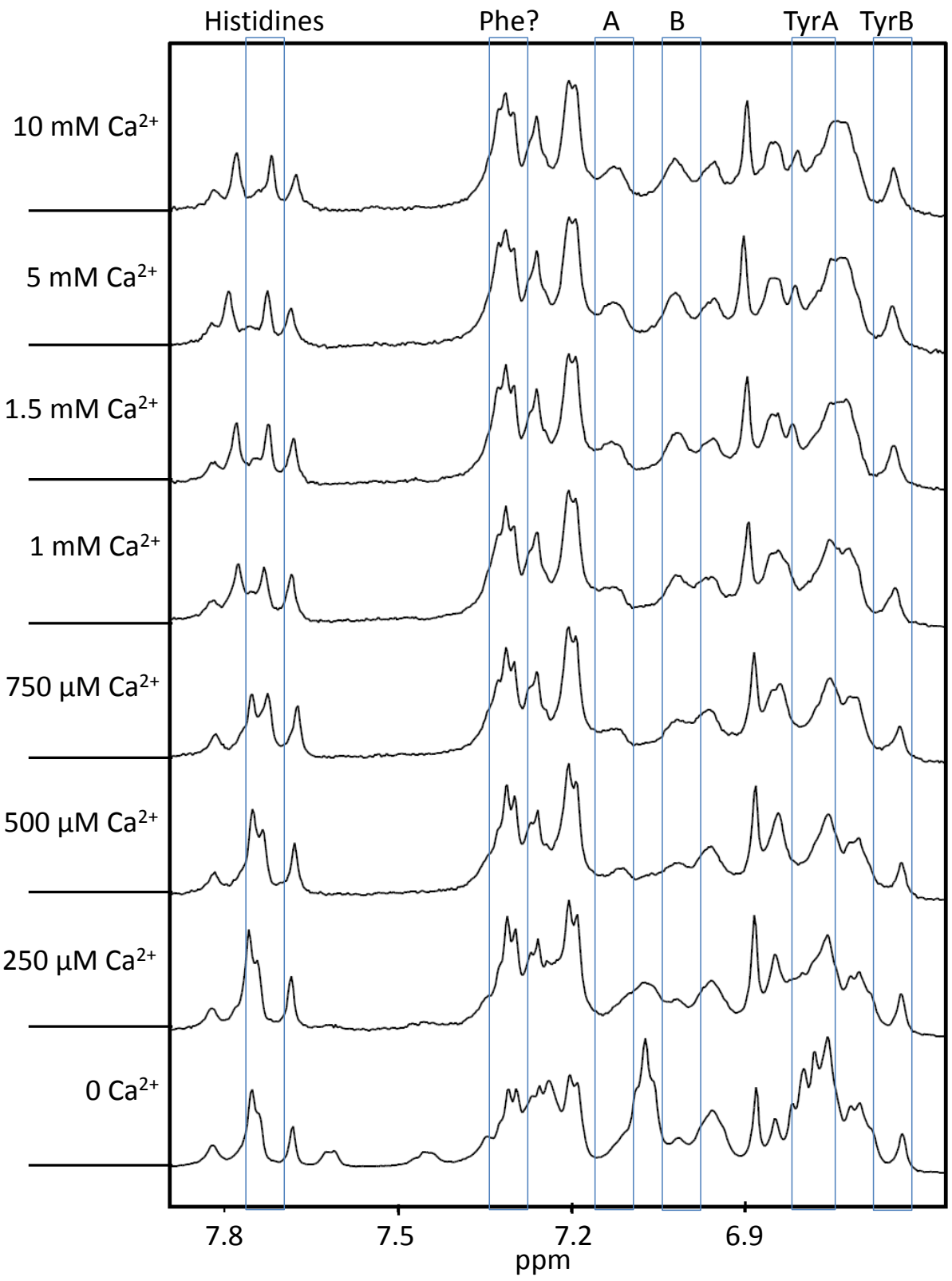


Figure 3.5 1D ^{1}H NMR Ca^{2+} titration of hN-1_{4-7} at pH 7.5 and *I* = 0.15. Histidine, phenylalanine and tyrosine peaks (A, B) are indicated (boxes). TyrA at around 6.8 ppm shifts up to 5 mM [Ca^{2+}] indicative of an intermediate affinity Ca^{2+} binding site. No obvious effect of Ca^{2+} on TyrB was observed. Boxes A and B contain further peaks affected by Ca^{2+}.

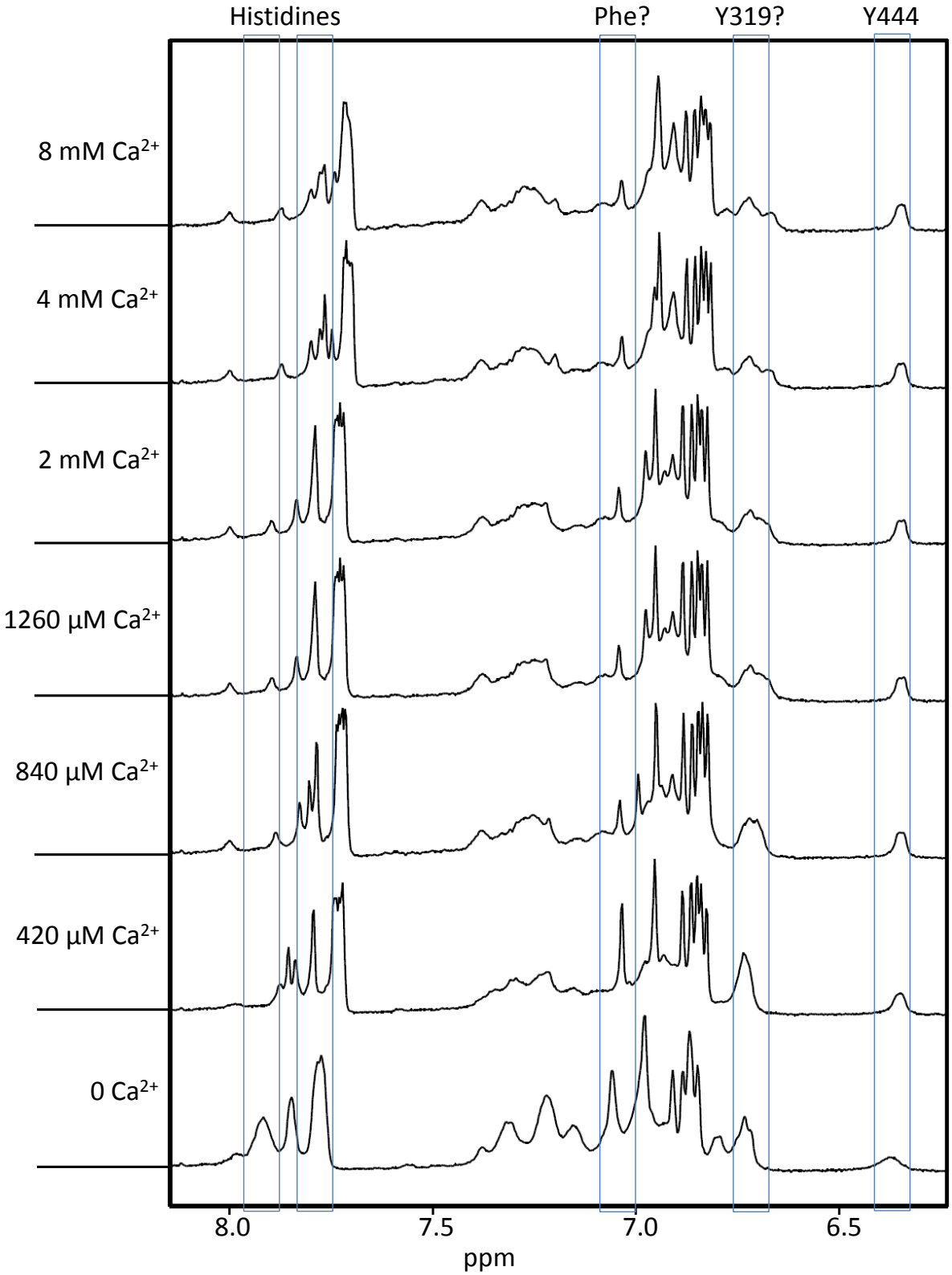


Figure 3.6 1D ^{1}H NMR Ca^{2+} titration of hN-$1_{8\text{-}11}$ at pH 7.5 and *I* = 0.15. Indicated are changes upon Ca^{2+} binding to peaks (boxes) in the histidine, phenylalanine and tyrosine region. Y444 of EGF11, which was previously assigned in hN-$1_{11\text{-}13}$, shows high Ca^{2+} affinity. A tyrosine at 6.7 ppm shifts until 4 mM Ca^{2+} and could represent a residue in EGF8 or EGF9. Further Ca^{2+} studies (chapter 4), however indicated that it is Y319 of a relatively high N-terminal Ca^{2+} affinity binding site (EGF 8).

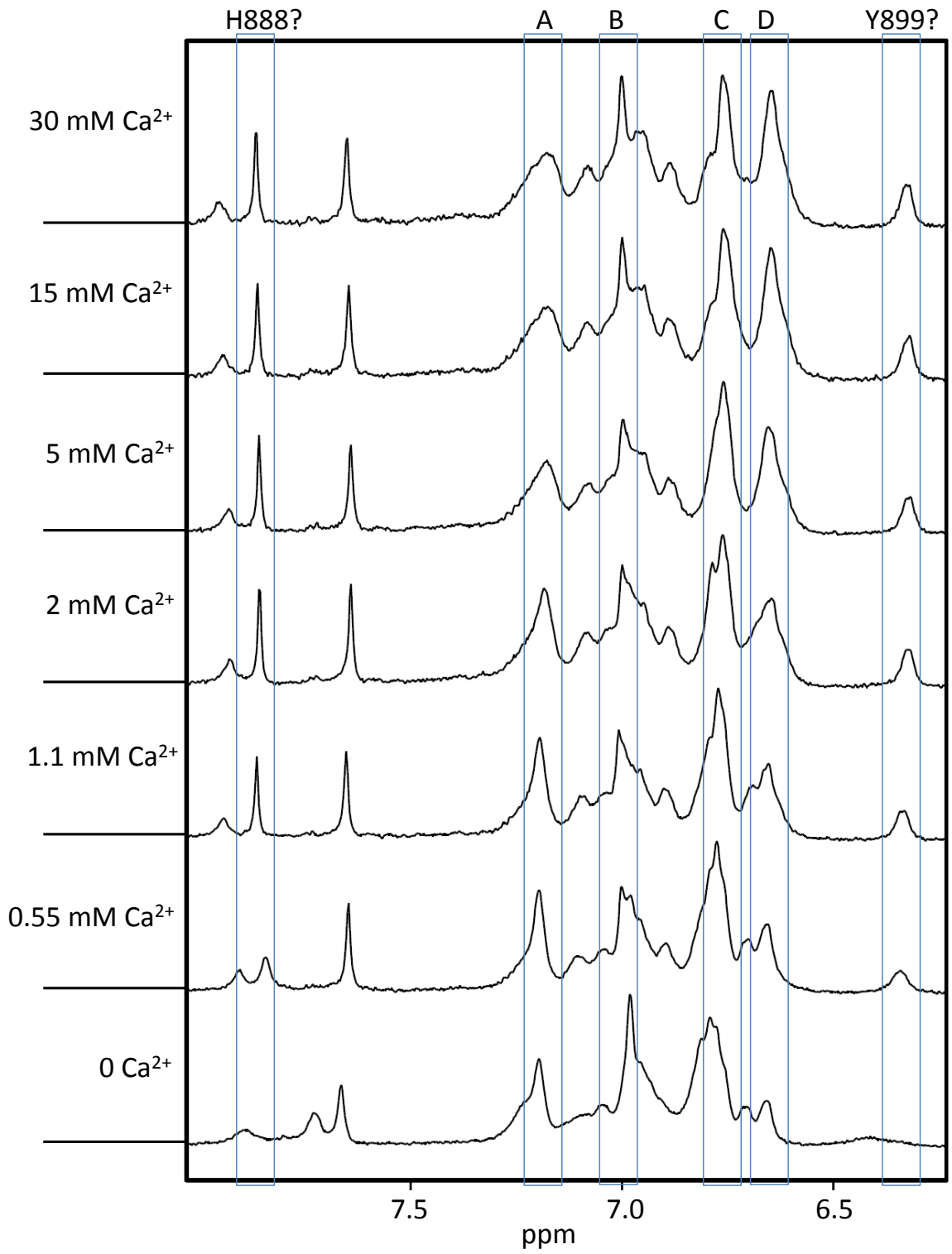


Figure 3.7 1D ^{1}H NMR Ca^{2+} titration of hN-1_{20-23} at pH 7.5 and *I* = 0.15. Boxes indicate the spectral changes upon Ca^{2+} additions. Similar to hN-1_{7-9}, Ca^{2+} has a huge impact on peak position, suggesting the presence of correctly folded cbEGF domains. Possible assignments for H888 and Y879 (EGF23) are presented.

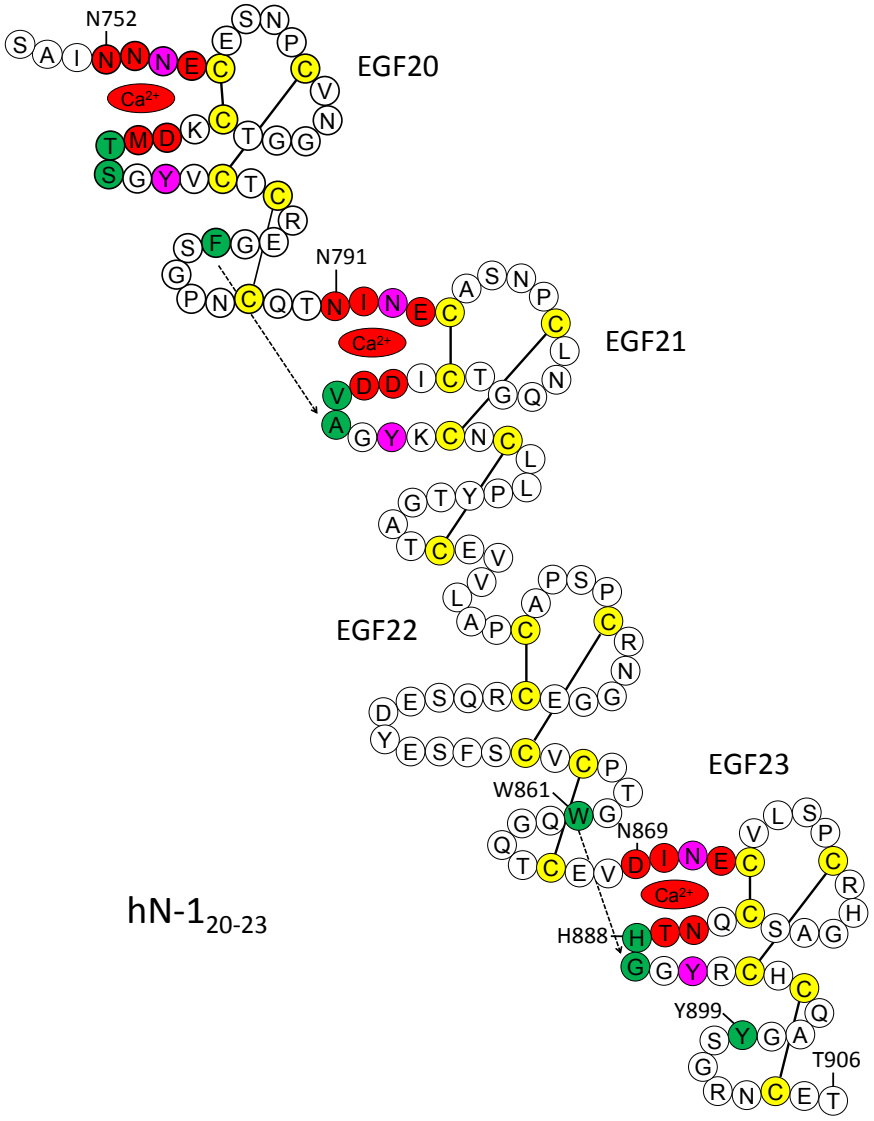


Figure 3.8 Schematic diagram of hN-1_{20-23}. Ca^{2+} and Ca^{2+} ligands are indicated in red, interdomain packing residues in green, cysteines in yellow and conserved residues for intradomain stability in purple. The arrows indicate regions that are involved in interdomain packing.

4 Ca^{2+} dissociation constants

4.1 Introduction

The extracellular region of human Notch-1 mainly consists of 36 epidermal growth factor-like (EGF) domains, many of which are calcium binding (cbEGF) (Figure 4.1).

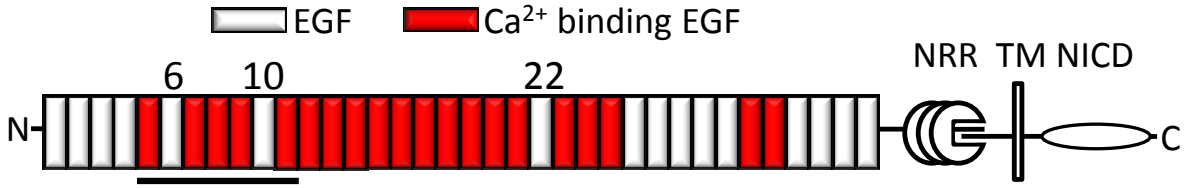


Figure 4.1 Notch-1 receptor organisation. The extracellular region of human Notch-1 consists of 36 EGF domains and the NRR region. The N- and C-terminus, EGF6, EGF10 and EGF22, the transmembrane (TM) domain and the intracellular domain (NICD) are indicated. In this study, Ca^{2+} affinities of EGF5-11 (black bar) are measured.

EGF domains are common modules in Notch and many other proteins and show a high degree of consensus homology (Bork et al., 1996). Within the last twenty years, structures of EGF domains from a wide range of proteins, such as factor IX, factor IXa, the LDL receptor, the Notch receptor and fibrillin, have been determined (Saha et al., 2001, Cordle et al., 2008a, Downing et al., 1996, Brandstetter et al., 1995, Rao et al., 1995). Whereas a consensus structure for individual EGF domains was identified, inter-domain tilt and twist angles between tandem EGF domains varied between extended and rigid such as in the Notch-$1_{11\text{-}13}$ crystal structure, or bent as in factor IXa (Figure 1.9 and Figure 1.15). Therefore, a detailed analysis of domain interactions, angles and flexibility is the prerequisite for understanding the cell surface organisation of the Notch receptor.

In EGF domains, a consensus sequence of D/N-x-D/N-E/Q-x_m-D/N*-x_n-Y/F (where * indicates possible β-hydroxylation, and m/n are variable) is predictive for Ca^{2+} binding (Rees et al., 1988, Handford et al., 1991, Mayhew et al., 1992). Interestingly, most Ca^{2+} binding proteins have seven ligands arranged on the vertices of a pentagonal bipyramid, whereas a minority shows six ligands in a distorted octahedral arrangement (Strynadka and James, 1989). A link

between Ca^{2+} binding (Figure 4.2) and interdomain rigidity has been suspected from early studies and is mainly due to hydrophobic interactions between a conserved aromatic residue (Y444) at the open end of the minor β-sheet and a hydrophobic residue (I471) at the end of the major β-hairpin of the following domain. Ca^{2+} measurements in the NH_2-terminal EGF domain of factor X revealed a 20-fold increase in affinity after the addition of a further 44 residues at the N-terminus (Valcarce et al., 1993). In the cbEGF32-33 domain pair of fibrillin-1, a 25-fold higher affinity was seen for the C-terminal (K_d = ~350 µM) than for the N-terminal domain (K_d = 9 mM), measured at physiological conditions of pH (7.5) and ionic strength (I = 0.15) (Knott et al., 1996).

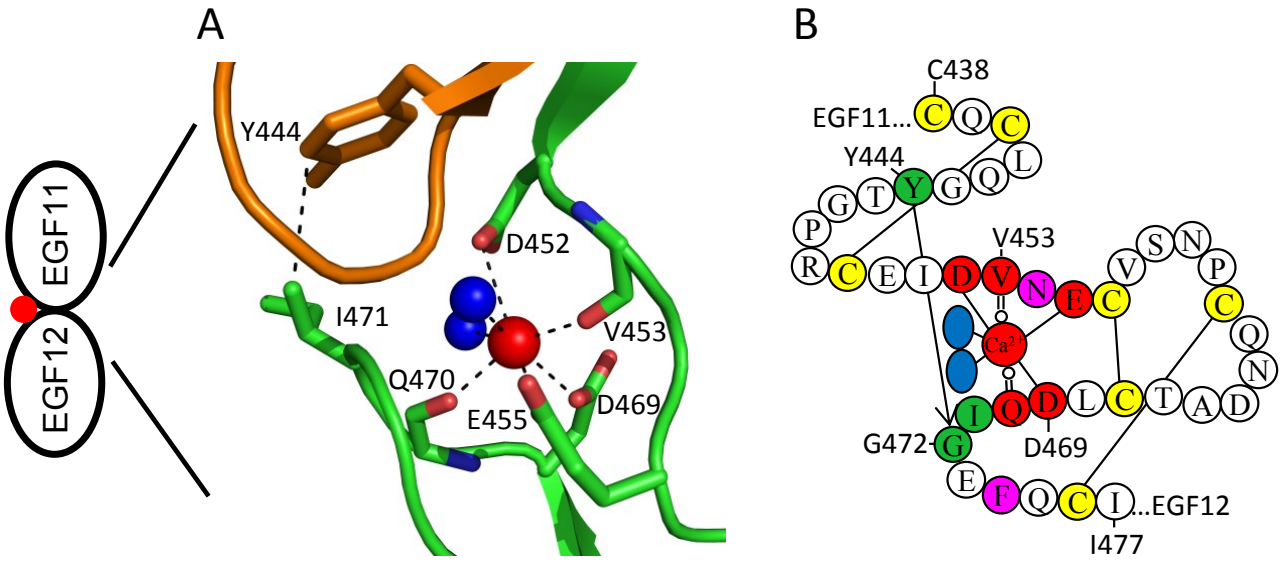


Figure 4.2 Ca^{2+} binding in EGF12 of human Notch-1. The crystal structure of hN-1_{11-13} shows seven oxygen ligands for Ca^{2+} (red) binding in EGF12 (A), two of which are oxygens from water molecules (blue). The ligands are arranged in a pentagonal bipyramid, the two vertices on the top/bottom are represented by one water molecule and the side chain of D469. The pentagonal base is formed by another water molecule, the backbone carbonyls from V453 and Q470, and side chain oxygens from D452 and E455. In (A), residues belonging to EGF11 are orange, and those belonging to EGF12 are green. (B) Ca^{2+} and ligands are shown in red, interdomain packing residues in green and cysteines in yellow. Note the conserved interdomain packing between Y444 of EGF11 and I471 in EGF12 which is predictive for high Ca^{2+} affinity and interdomain rigidity (arrow). F474 (purple) is a conserved aromatic residue in EGF12 that stabilizes the domain structure without direct involvement in Ca^{2+} binding. N454 (purple) forms an ASX motif with C465. G472 is a conserved glycine that forms a turn in the β–hairpin and contributes to the interdomain packing with its Hα. G472 is one residue after the hydrophobic packing residue, often referred to as the XG motif and also involved in packing.

Similarly, the affinity of isolated cbEGF13 of fibrillin-1 (K_d = 2-3 mM) is enhanced more than 70-fold by the addition of cbEGF12 (Smallridge et al., 1999). The solution structure of fibrillin-1

cbEGF32-33 revealed a rigid, rod-like organisation, stabilised by Ca^{2+}. Further evidence for a rigid interface came from ^{15}N relaxation NMR studies, demonstrating that residues close to the Ca^{2+} binding site are more stable than the remaining structure (Werner et al., 2000, Smallridge et al., 2003). Various solution and X-ray structures confirmed that Ca^{2+} binding is predictive for a rod-like organisation for tandem cbEGF arrays (Downing et al., 1996, Smallridge et al., 2003, Hambleton et al., 2004, Cordle et al., 2008a).

Similar findings were reported for heterologous domain pairs comprising transforming growth factor β-binding protein-like (TB) domains located N-terminal of cbEGF domains in fibrillin-1. A loose interface at the TB6-cbEGF32 junction of fibrillin-1 was suggested by only slight enhancement of the Ca^{2+} affinity of cbEGF32 from K_d = 4.3 mM, in cbEGF32-33, to K_d = 1.6 mM, in TB6-cbEGF32 (Kettle et al., 1999). In contrast, the Ca^{2+} affinity of cbEGF23 in TB4-cbEGF23 was 16 nM, five orders of magnitude higher than cbEGF32 in TB6-cbEGF32. The observation of Ca^{2+} dependent interdomain NOE signals between the aromatic packing residues of domain swap mutant TB4-cbEGF32 with engineered packing residues, but not for TB6-cbEGF32 confirmed the hypothesis that interdomain rigidity based on hydrophobic contacts and high affinity Ca^{2+} binding is linked and showed that this effect is not limited to cbEGF domain pairs (Jensen et al., 2005).

Rotary shadowing of a 12 EGF domain fragment of human fibrillin-1, a protein rich in multiple tandem cbEGF domains, showed a more extended and ordered structure in the presence of calcium compared to that observed in EDTA (Reinhardt et al., 1997). Similarly, microfibrils, an ultrastructure consisting mainly of fibrillin-1 monomers, showed more flexibility in the absence of Ca^{2+} (Cardy and Handford, 1998).

Table 4.1 summarises K_d values obtained for various cbEGF domain constructs. K_d values for cbEGF with a preceding N-terminal EGF domain range from 0.3 µM in cbEGF31 to ~350 µM in cbEGF33.

Table 4.1 Summary of published K_d values of cbEGF domains in different proteins, measured at physiological pH (7.5) and ionic strength (I = 0.15).

protein	construct	domain	K_d	published as
fib-1	cbEGF12-13	cbEGF12/13	1.6 mM /<30 µM	Smallridge et al., 1999
fib-1	cbEGF32-33	cbEGF32/33	~9 mM / ~350 µM	Knott et al., 1996
fib-1	cbEGF13-14	cbEGF13/14	~3 mM / <100 µM	Whiteman et al., 1998
fib-1	cbEGF29-31	cbEGF29/30/31	~1.5 mM / 3.7 µM / 0.3 µM	Suk et al., 2004
CRB1	EGF18-cbEGF19	cbEGF19	220 µM	Davis et al., 2007
hN-1	EGF12-13	EGF13	34 µM	Rand et al., 1997

In this chapter, Ca^{2+} affinities of EGF domains in the EGF5-11 region of human Notch-1 are measured. This region is immediately N-terminal to the ligand-binding domain EGF11-12 (Rebay et al., 1993) and comprises non calcium-binding EGF6 and EGF10.

All cbEGF domains in this region show the consensus Ca^{2+} binding sequence and the aromatic packing residues in the preceding domains (Table 1.1, section 1.4.4). This study provides an initial insight into which interfaces are likely to be flexible and rigid (Figure 4.3) on the cell surface and is thus a prerequisite for a detailed structural analysis and modelling of the N-terminal region of Notch under physiological conditions.

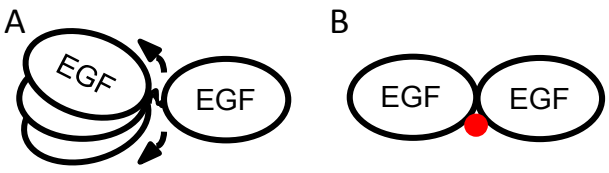


Figure 4.3 EGF interdomain interactions based on Ca^{2+} binding. The study of affinities within the region EGF5-11 of human Notch-1 may give insight into which of the studied domains are flexible (A in the absence of Ca^{2+} or with low affinity) or rigid (B). Ca^{2+} is indicated in red. Note that Ca^{2+} in the N-terminal domain was not depicted for clarity.

4.2 Determination of Ca^{2+} affinities

4.2.1 Binding models and affinity

Binding of molecules to each other can be described in its simplest case in which molecules of type A and molecules of type B form complexes, AB, in a reversible reaction, with the corresponding rates k_{on} and k_{off} (equation 1).

(1) $$A + B \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} AB$$

Equilibrium is reached when concentrations no longer change; the dissociation constant (K_d) for the AB complex at equilibrium is defined as (2).

(2) $$K_d = k_{off} / k_{on} = [A][B] / [AB]$$

The dissociation constant is easy to calculate if the concentration of each compound is measurable. However, this is experimentally not straightforward and usually achieved in a titration with increasing concentration of one of the compounds. For Ca^{2+} binding to cbEGF domains (P for protein), the equation can be rewritten (3).

(3) $$K_d = [Ca^{2+}_{free}]*[P_{free}]/[Ca^{2+}P]$$

The K_d is also defined as $[Ca^{2+}_{free}]$ when half of the protein has calcium bound (4).

(4) $$\text{For } [P_{free}] = [Ca^{2+}P];\ K_d = [Ca^{2+}_{free}]$$

4.2.2 NMR in measurement of low Ca^{2+} affinities

Ca^{2+} affinities in cbEGF domains have been measured previously by NMR (Persson et al., 1989, Smallridge et al., 1999, Knott et al., 1996, Suk et al., 2004, Jensen et al., 2005, Whiteman et al., 1998, Kettle et al., 1999, Davis et al., 2007), equilibrium dialysis (Glanville et al., 1994), intrinsic fluorescence (Astermark et al., 1992, Malby et al., 2001) and chromophoric chelator

methods (Rand et al., 1997, Suk et al., 2004, Jensen et al., 2005, Schneider et al., 2010, Whiteman et al., 2007). Of these methods, only NMR is able to describe Ca^{2+} binding to specific domains in tandem cbEGF domain constructs. In a standard binding titration, Ca^{2+} is added step by step to a protein solution. Ca^{2+} binding then leads to a change in the chemical environment of certain residues and to a change in the chemical shift in the spectrum. The exchange rate ($k_{ex} = k_{on} + k_{off}$) and the chemical shift difference between the bound and unbound states determine whether peaks are observed to shift (fast exchange) or whether peak intensity changes (slow exchange) are observed during the titration (Figure 4.4).

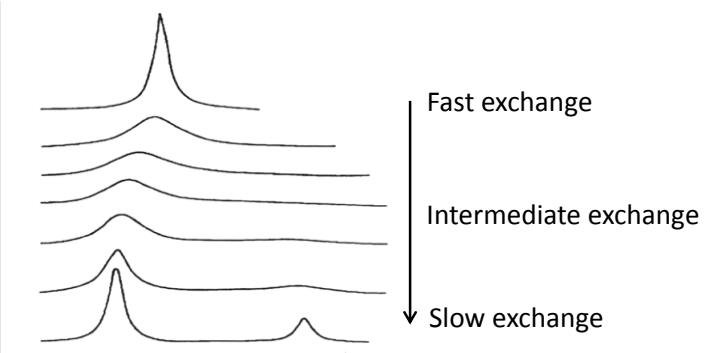


Figure 4.4 The effect of chemical exchange. The exchange rate k_{ex} indicates how many times per second a nucleus can interconvert between state A and B. If k_{ex} is larger than the resonance frequency difference between nuclei in state A and B, an average over the two states is seen in the spectrum (fast exchange). Line broadening occurs if exchange rate and frequency differences are similar (intermediate exchange). For very slow exchange, two separate peaks can be seen.

K_d values are most accurately determined by NMR in the fast exchange limit. Here, the ratio of the chemical shift at each titration point and the chemical shift at saturation correlates directly with the fraction of Ca^{2+}-bound protein. In slow exchange, K_d values can be determined by following the change of peak intensities. However, slow exchange is often accompanied by line broadening. Therefore, the accurate measurement of high affinity Ca^{2+} domains becomes more difficult by NMR. Note that frequency differences are smaller at lower magnetic field and exchange rates higher at high temperature. Furthermore, the chemical shift difference between different states correlates with the difference of the chemical environment. Therefore, by

avoiding resonances with chemical shift differences close to the exchange rate, or by alternating temperature and magnetic field strength, systems can in certain cases be shifted towards fast exchange and K_d values be determined with higher accuracy.

In fast exchange, the Ca^{2+} dissociation constant is calculated by fitting to equation (5), with ($\Delta = \delta - \delta_{[Ca2+]=0}$ and $\Delta_0 = \delta_{Sat} - \delta_{[Ca2+]=0}$) using the software Kaleidagraph 3.5 (δ_{Sat} is the chemical shift at saturation).

(5) $$\Delta/\Delta_0 = [Ca^{2+}_{free}] / (K_d + [Ca^{2+}_{free}])$$

NMR titrations were performed under physiological conditions; at pH 7.5 in 5 mM Tris and 150 mM NaCl (I = 0.15). Two-dimensional ^{1}H NOESY NMR has been used widely to determine K_d values by following the Ca^{2+} dependent chemical shifts of two highly conserved aromatic residues (Jensen et al., 2005, Handford et al., 1995, Whiteman et al., 2007). The first residue is a conserved aromatic residue two residues before the fourth conserved cysteine (Figure 4.2) that stabilises the domain fold (Figure 1.11) and measures Ca^{2+} binding within that domain. The second residue is a conserved aromatic residue, four residues after the fifth conserved cysteine of the preceding domain that enables hydrophobic packing with the end of the major β-hairpin in the following cbEGF domain and measures Ca^{2+} binding in the latter (Figure 4.2). The observed chemical shifts of these protons are located in an isolated region of a proton spectrum at around 6-8 ppm. The low number of aromatic residues, generally seen in these constructs facilitates assignment.

4.2.3 Chromophoric chelation to measure high Ca^{2+} affinities

The affinity of Ca^{2+} sites with a low K_d can be measured accurately by chromophoric chelation, which has been used to measure K_d values in proteins, such as protein S, calbindin, fibrillin and Notch (Linse et al., 1991, Dahlbäck et al., 1990, Berggård et al., 2002, Jensen et al.,

2005, Rand et al., 1997). Ca^{2+} is titrated into a solution containing a chromophoric chelator with a Ca^{2+} binding site. Ca^{2+} binding leads to a change in the conformation and hence absorbance spectrum of the chelator; the free Ca^{2+} concentration can then be determined. The addition of a Ca^{2+} binding protein results in a competition for Ca^{2+} binding with the chelator, clearly visible as the change in the degree in absorbance (Figure 4.5).

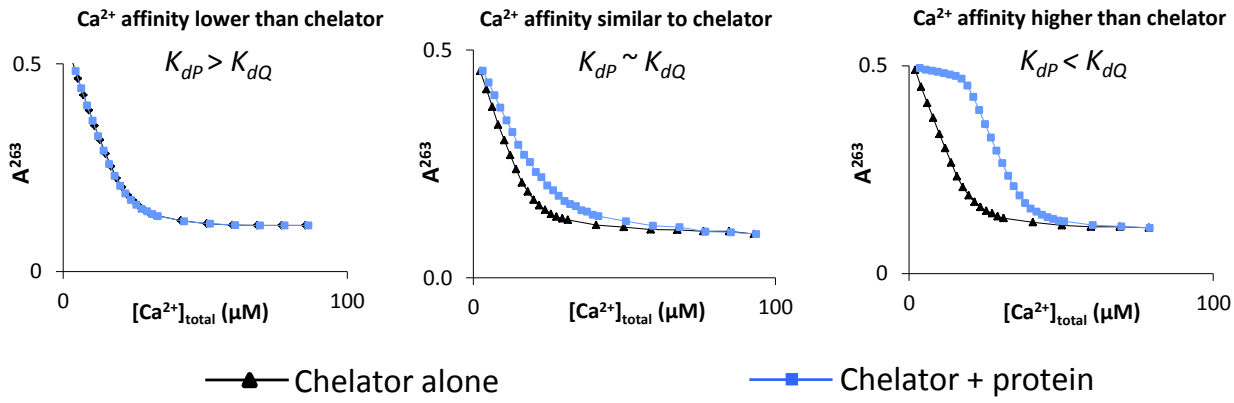


Figure 4.5 Illustration of chromophoric chelation. No competition is observable if the K_d of the Ca^{2+} binding protein (P) is much higher than the K_d of the chelator (Q), ($K_{dP} > K_{dQ}$). Competition occurs with similar affinities ($K_{dP} \sim K_{dQ}$) or higher affinities ($K_{dP} < K_{dQ}$). (Figure adapted from Jemima Cordle - DPhil thesis 2005).

Experiments are most accurate when the protein has a K_d similar to that of the chelator. Here, the chromophoric chelator 5,5-Br_2BAPTA which has a K_d for Ca^{2+} of 1.6 μM is used. Note that a chelator with a higher K_d (5-NBAPTA) has been used previously (Rand et al., 1997), but is no longer commercially available.

4.3 Calcium binding in hN-$1_{4\text{-}7}$ and hN-$1_{5\text{-}7}$

2D NOESY ^{1}H NMR Ca^{2+} titrations of EGF5 and EGF7 were first performed for 300 μM hN-$1_{5\text{-}7}$ and subsequently for 500 μM hN-$1_{4\text{-}7}$, (pH 7.5, I = 1.5) (Figure 4.6 and Figure 4.7). In hN-$1_{4\text{-}7}$, EGF5 is in a native context since it is sandwiched by two EGF domains. Ca^{2+} binding in EGF7 was analysed by following the chemical shift of the aromatic consensus residue in EGF7, Y279, and the aromatic packing residue F249 within EGF6. The chemical shift of the aromatic

markers showed no further change above an equimolar concentration of Ca^{2+}, revealing high Ca^{2+} affinity in EGF7 for hN-1_{5-7} (Figure 4.7 A and B) and hN-1_{4-7} (Figure 4.7 C and D).

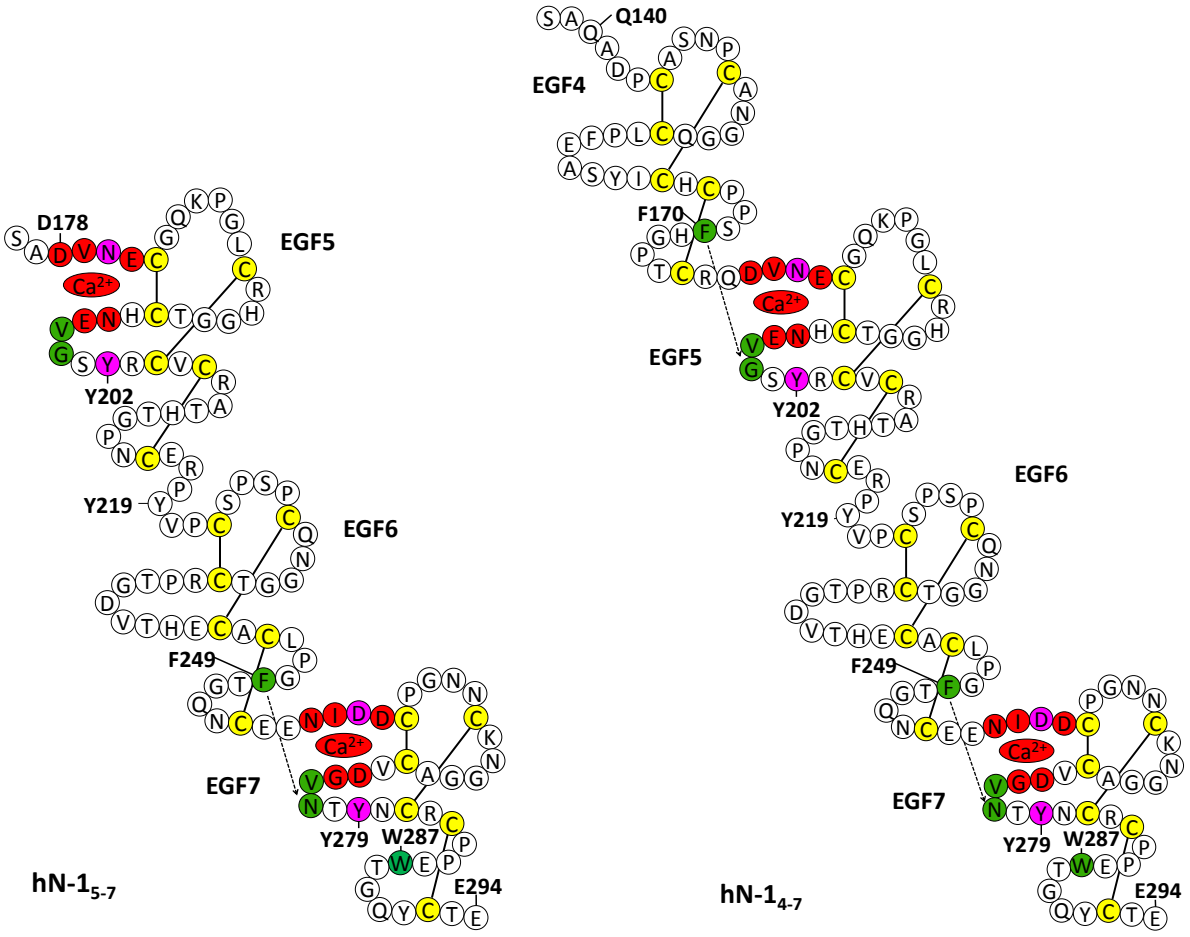


Figure 4.6 Schematic diagrams of hN-1_{5-7} (left) and hN-1_{4-7} (right). Aromatic markers are indicated, calcium and ligands in red, packing residues in green, conserved residues for intradomain stability in purple, and cysteines in yellow. Interdomain packing is indicated (arrow).

Further addition of Ca^{2+} up to 25 mM, revealed a low affinity Ca^{2+} binding site in EGF5 of hN-1_{5-7} as the aromatic consensus residue of EGF5, Y202, shifted with increasing amounts of Ca^{2+}, an example of fast exchange (B). In hN-1_{4-7}, Ca^{2+} binding to EGF5 was analysed following the chemical shift of F170 (EGF4), since Y202 overlapped with other resonances (D). The N-terminal linkage of EGF4 was shown to increase the Ca^{2+} affinity of EGF5, since saturation was observed at 5 mM Ca^{2+}.

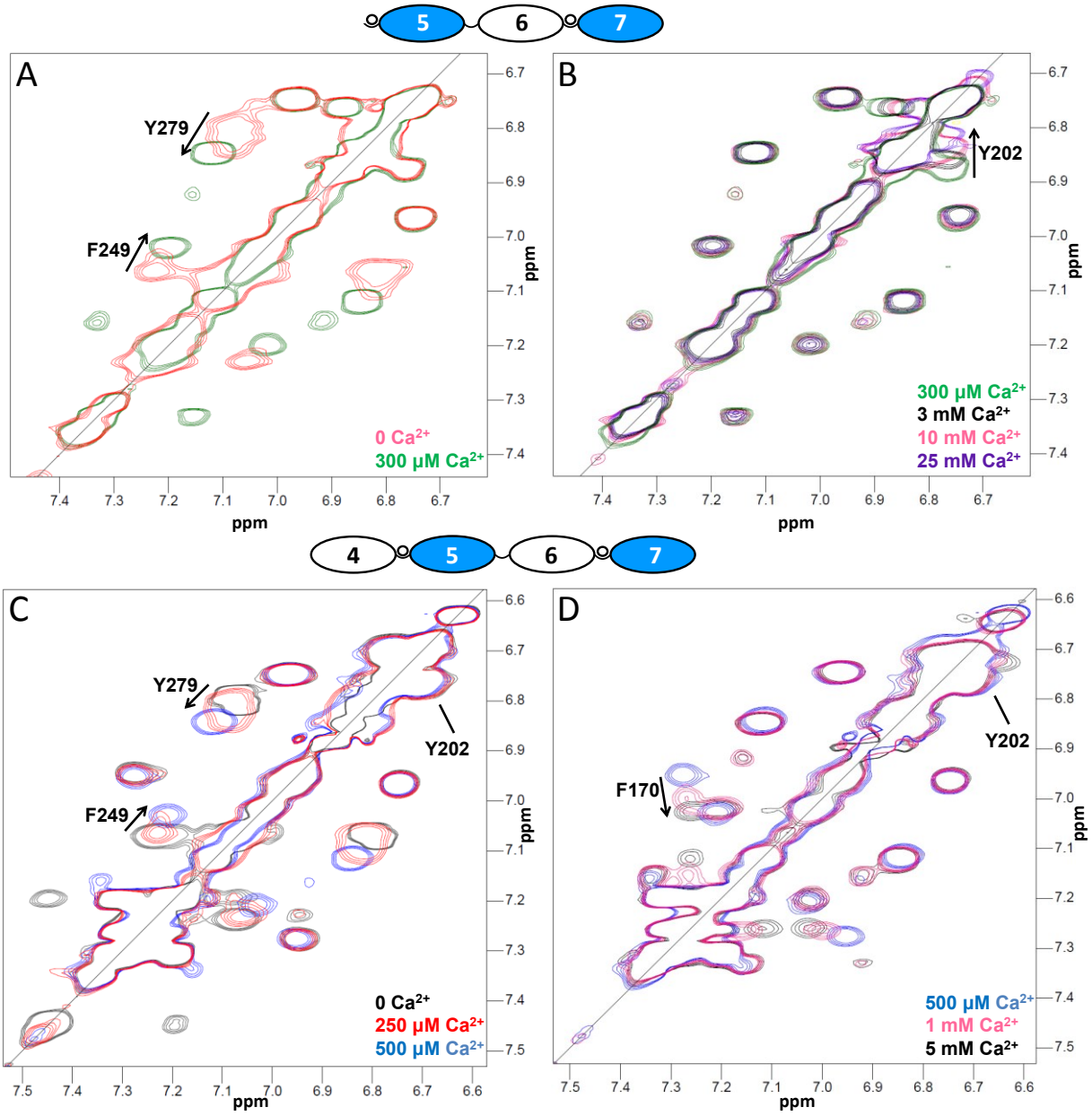


Figure 4.7 2D NOESY Ca^{2+} titration of the aromatic region in hN-$1_{5\text{-}7}$ (A-B) and hN-$1_{4\text{-}7}$ (C-D) at pH 7.5 and Ionic strength *I* = 0.15. Schematic diagrams of both constructs are shown above the spectra. Observation of the maximum shift change for Y279 and aromatic packing residue F249 at an equimolar concentration of Ca^{2+} in hN-$1_{5\text{-}7}$ (A) and hN-$1_{4\text{-}7}$ (C) revealed high Ca^{2+} affinity in EGF7. Further addition of Ca^{2+} (B) showed fast exchange of Y202 in hN$1_{5\text{-}7,}$ revealing low Ca^{2+} affinity in EGF5. Note that Y202 showed overlap in the spectrum of hN-$1_{4\text{-}7}$, and Ca^{2+} binding to EGF5 was analysed by following the chemical shift of F170 (D). The affinity of EGF5 is increased by the N-terminal linkage of EGF4, as no further shift changes were observed above 5 mM Ca^{2+} (not shown). Arrows indicate direction of chemical shifts during the titration.

The calculation of the dissociation constant for the high affinity site in EGF7 was based on Ca^{2+} chelation experiments with the chromophoric chelator 5,5-Br_2BAPTA (Figure 4.8).

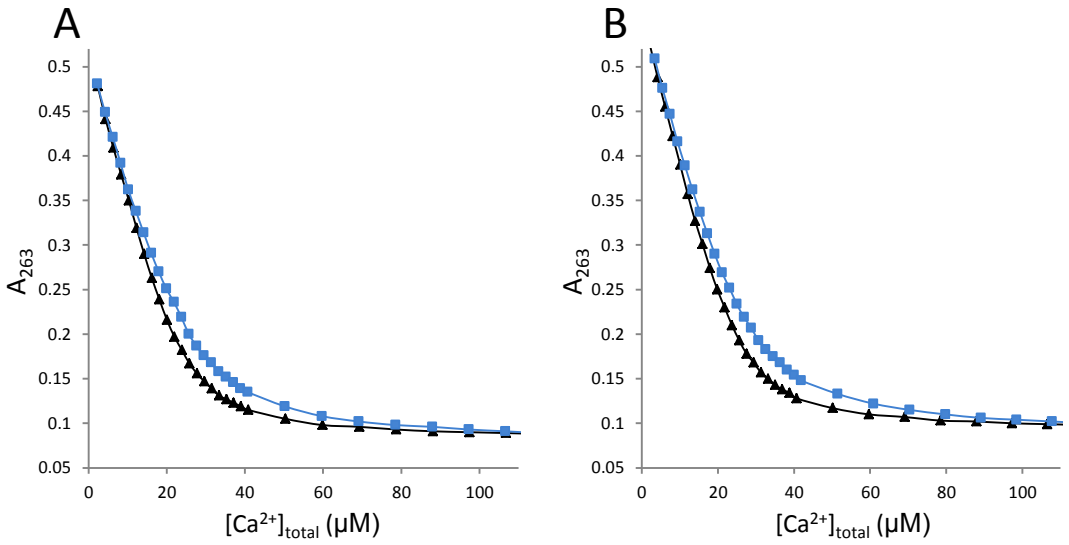


Figure 4.8 Ca^{2+} titration curve of 5,5-Br_2BAPTA with hN-1_{5-7} (A) and hN-1_{4-7} (B) to measure Ca^{2+} affinities in EGF7, at pH 7.5 and *I* = 0.15. Ca^{2+} titration in the absence of protein is shown by black triangles. Addition of hN-1_{5-7} (blue squares in A) and hN-1_{4-7} (blue squares in B) results in competition for Ca^{2+} binding and a reduction in the slope of the binding curve. Similar levels of competition with the chelator suggests structural identity at the C-terminus of both constructs and no effect of EGF5 on the competition.

The hN-1_{5-7} and hN-1_{4-7} datasets showed little competition suggesting that the K_d for EGF7 is significantly higher than 1.6 μM. Fitting of the hN-1_{5-7} dataset resulted in a K_d of 16 ± 8 μM for EGF7 (n = 3, independent protein preparations). The lack of accuracy in these measurements is due to the fact that EGF7 doesn't compete effectively with the chelator. One-site fitting of the hN-1_{4-7} datasets with 4 independent repeats (n =4, same protein preparation) resulted in a K_d of 27 ± 3 μM. The error for the fitting was slightly reduced assuming two Ca^{2+} binding sites, resulting in a K_d of 24 ± 4 μM for EGF7. The K_d values for EGF5 in the two-site fitting, on the other hand, varied between 30 μM and 300 μM; this variation is due to the weak affinity in EGF5 compared to the chelator. Although the error of the Ca^{2+} affinities for EGF7 in both constructs was high, the K_d values of 24 ± 4 μM (hN-1_{4-7}) and 16 ± 8 μM (hN-1_{5-7}) were in line with the assumption that the addition of EGF4 to hN-1_{5-7} has no effect on Ca^{2+} binding in EGF7.

Ca^{2+} affinities in EGF5 were fitted following the chemical shift change of resonances Y202 in hN-1_{5-7} (Figure 4.9 A) and F170 in hN-1_{4-7} (Figure 4.9 B).

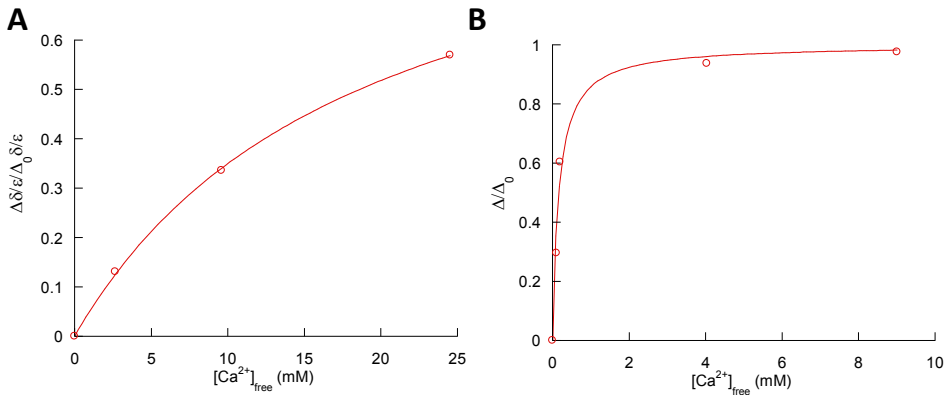


Figure 4.9 Plots of chemical shift change versus free Ca^{2+} in hN-1_{5-7} (A) and hN-1_{4-7} (B). Markers were Y202 in hN-1_{5-7} and F170 in hN-1_{4-7}. In A, the K_d of EGF5 for Ca^{2+} was fitted to 19 mM. In B, the K_d of EGF5 for Ca^{2+} was fitted to 170 µM. Ca^{2+} dissociation constants were derived as described previously (Jensen et al., 2005).

In both cases, $[Ca^{2+}]_{free}$ was calculated assuming that the Ca^{2+} binding site in EGF7 is fully saturated during the titration. A K_d of 19 mM was calculated for EGF5 in hN-1_{5-7}. The addition of EGF4 increased the affinity of EGF5 for Ca^{2+} by 2 orders of magnitude to 170 µM in hN-1_{4-7}. Figure 4.10 shows a summary of K_d values for Ca^{2+} binding in hN-1_{5-7} and hN-1_{4-7}. All K_d values lie well in the range of previous reports (Table 4.1) and demonstrate the usefulness of NMR and chromophoric chelation in the measurement of Ca^{2+} affinities in EGF domains. This finding is in line with previous reports of an increased affinity by the addition of an N-terminal domain.

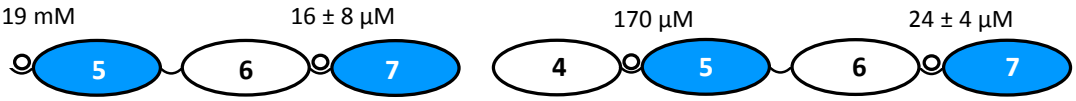


Figure 4.10 Ca^{2+} affinities in constructs hN-1_{5-7} (left) and hN-1_{4-7} (right). Affinity measurements confirm a reasonably high Ca^{2+} affinity in EGF7 and a lower Ca^{2+} affinity in EGF5. Addition of EGF4, as expected, increased the affinity of EGF5 by around two orders of magnitude. Both sites would be expected to be fully saturated under physiological conditions.

4.4 Calcium binding in hN-1_{9-11} and hN-1_{8-11}

2D NOESY ^{1}H NMR Ca^{2+} titrations were performed for 300 µM hN-1_{9-11} and 420 µM hN-1_{8-11} (Figure 4.11 and Figure 4.12).

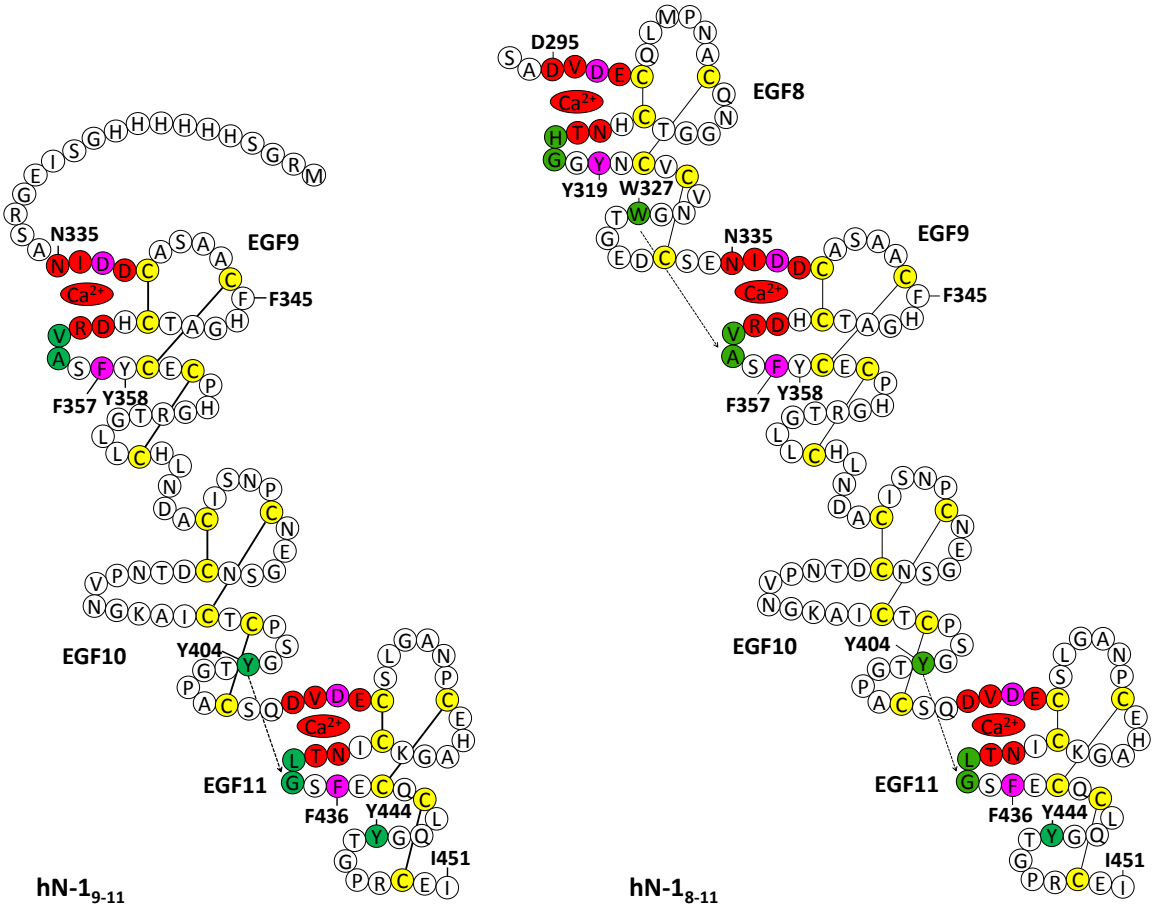


Figure 4.11 Schematic diagrams of hN-1_{9-11} (left) and hN-1_{8-11} (right). Aromatic markers are indicated, calcium and ligands in red, packing residues in green, conserved residues for intradomain stability in purple, and cysteines in yellow. Interdomain packing is indicated (arrow). Because of a cryptic cleavage site for factor Xa in EGF9, the his-tag was not removed in hN-1_{9-11}. In, hN-1_{8-11}, the his-tag was cleaved off after engineering an enterokinase cleavage site in between the his-tag and EGF8.

In hN-1_{9-11}, resonances of the aromatic consensus residue in EGF11, F436, and packing residue, Y404 (EGF10), appear at 300 µM Ca^{2+} (Figure 4.12 A). No further change of these resonances upon Ca^{2+} addition (B) suggests high Ca^{2+} affinity in EGF11. However, further addition of Ca^{2+} shows shifts for the aromatic consensus residue in EGF9, F357 revealing low Ca^{2+} affinity in EGF9. The addition of EGF8 in hN-1_{8-11} results in a significant increase in the Ca^{2+} affinity of EGF9 as EGF9 (F357) and EGF11 (F436) compete for binding (C) and both domains show saturation at 840 µM Ca^{2+}.

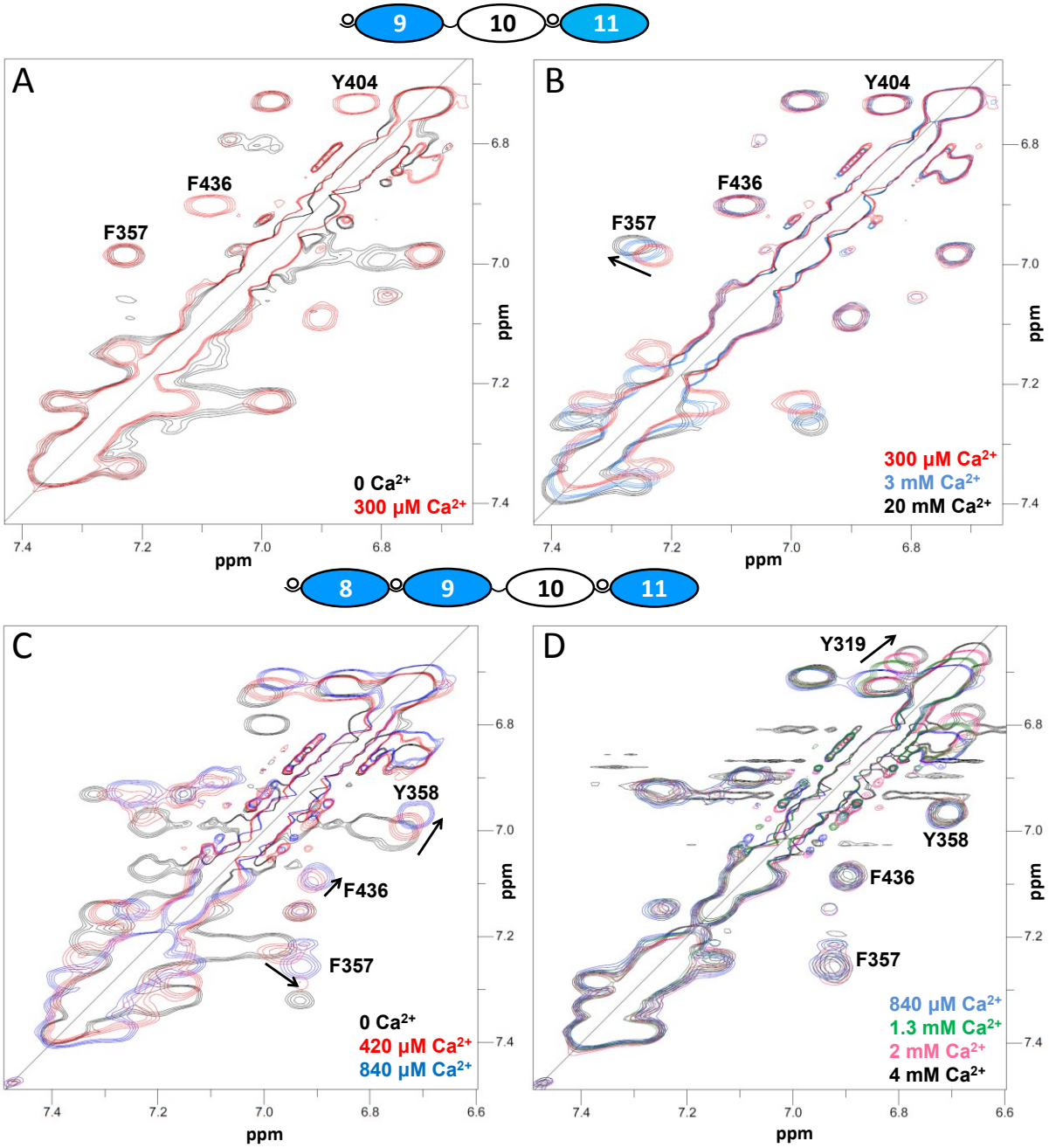


Figure 4.12 2D NOESY calcium titration of 300 µM hN-$1_{9\text{-}11}$ (A and B), and 420 µM of hN-$1_{8\text{-}11}$ (C and D) at pH 7.5 and *I* = 0.15. Schematic diagrams of both constructs are shown above the spectra. In A, the peak for F436 shows no change after addition of an equimolar concentration of Ca^{2+}, revealing high affinity for EGF11. In B, F357 shifts as more Ca^{2+} is added, revealing a low Ca^{2+} affinity binding site in EGF9 of hN-$1_{9\text{-}11}$. In C, the addition of EGF8 results in a significant increase in the Ca^{2+} affinity for EGF9 as EGF9 (F357) competes with EGF11 (F 436) for Ca^{2+} binding. The chemical shift of Y358, a non-conserved aromatic residue in EGF9, is shown as well. In D, Y319 of EGF8 shifts at higher concentrations of Ca^{2+}, revealing low Ca^{2+} affinity in EGF8. Arrows indicate direction of shifts during the titration.

In D, aromatic consensus residue Y319 of EGF8 shifts with fast exchange revealing low Ca^{2+} affinity in EGF8.

The dissociation constants for the high affinity sites in EGF11 and EGF9 were determined, based on Ca^{2+} chelation experiments with the chromophoric chelator 5,5-Br_2BAPTA. Fitting of the hN-1_{9-11} (Figure 4.13) dataset resulted in a K_d of 5.6 ± 0.6 µM for EGF11 (n = 3, independent protein preparations). Two-site fitting in hN-1_{8-11} showed affinities of 4.9 ± 0.2 µM and 26 ± 6 µM (n = 3, independent protein preparations). The similarity of K_d 4.9 ± 0.2 µM in hN-1_{8-11} and 5.6 ± 0.6 µM of EGF11 in hN-1_{9-11} suggests that the K_d of 4.9 ± 0.2 µM can be attributed to EGF11 and 26 ± 6 µM to EGF9.

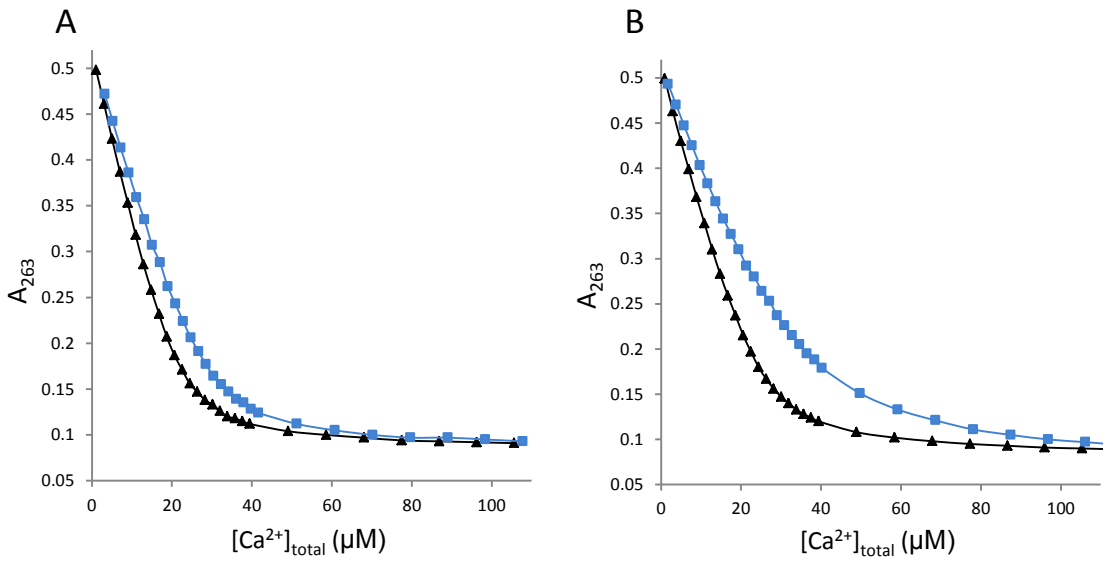


Figure 4.13 Ca^{2+} titration curve of 5,5-Br_2BAPTA and hN-1_{9-11} (A) and hN-1_{8-11}(B) at pH 7.5 and *I* = 0.15. Ca^{2+} titration without protein (black triangles) shows the changes in absorbance upon Ca^{2+} titration. Addition of hN-1_{9-11} (blue squares in A) and hN-1_{8-11} (blue squares in B) resulted in competition for Ca^{2+} binding and a reduction in the slope of binding curve. Note that significantly less competition is observed in hN-1_{9-11} than in hN-1_{8-11} confirming high Ca^{2+} affinity in EGF9 of hN-1_{8-11}. Ca^{2+} dissociation constants were derived as described previously (Jensen et al., 2005).

Affinities for N-terminal domains were calculated as previously, based on NMR chemical shift differences of F345 in hN-1_{9-11} and Y319 in hN-1_{8-11} (Figure 4.12 and Figure 4.14). The K_d of EGF9 in hN-1_{9-11} was fitted to 2.2 mM, the K_d of EGF8 in hN-1_{8-11} was fitted to 240 µM.

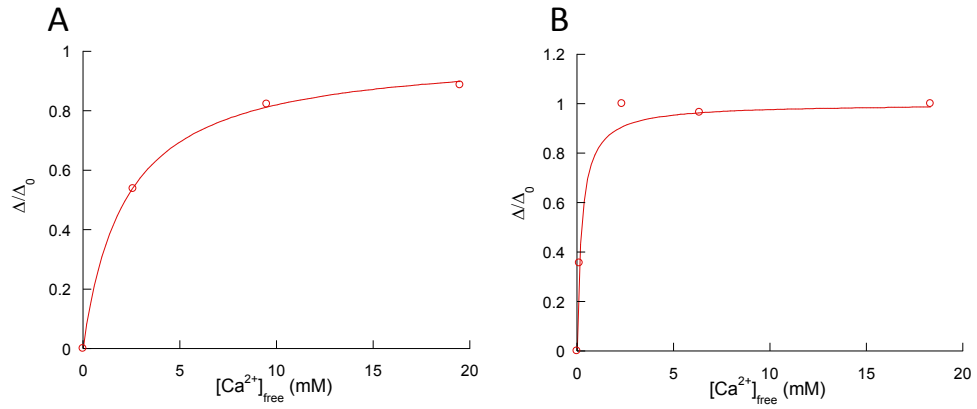


Figure 4.14 Changes in chemical shifts against free Ca^{2+} in hN-1_{9-11} (A) and hN-1_{8-11} (B). K_d values were 2.2 mM for EGF9 in hN-1_{9-11} and 240 µM for cbEGF8 in hN-1_{8-11}. Ca^{2+} dissociation constants were derived as described previously (Jensen et al., 2005).

Figure 4.15 shows an overview of measured affinities. The affinity of EGF9 was increased by around 2 orders of magnitude upon N-terminal linkage of EGF8, similar to the difference observed between hN-1_{5-7} and hN-1_{4-7}.

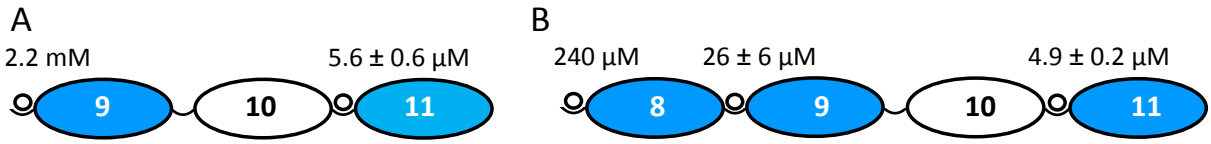


Figure 4.15 Ca^{2+} dissociation constants for constructs hN-1_{9-11} (A) and hN-1_{8-11} (B). Note that the affinity for EGF11 was similar in both constructs, whereas the affinity for the N-terminal domain in hN-1_{8-11} (EGF8) was around one order of magnitude higher than in hN-1_{9-11} (EGF9). The affinity of EGF9 for Ca^{2+} was two orders of magnitude higher in hN-1_{8-11} than in hN-1_{9-11}. The calcium binding sites would be expected to be saturated under physiological conditions.

4.5 Calcium binding in hN-1_{7-9}

Three Ca^{2+} binding sites were observed in 2D NOESY ^{1}H NMR Ca^{2+} titrations of hN-1_{7-9} (schematic diagram in Figure 4.16 and 2D NOESY Ca^{2+} titration in Figure 4.17).

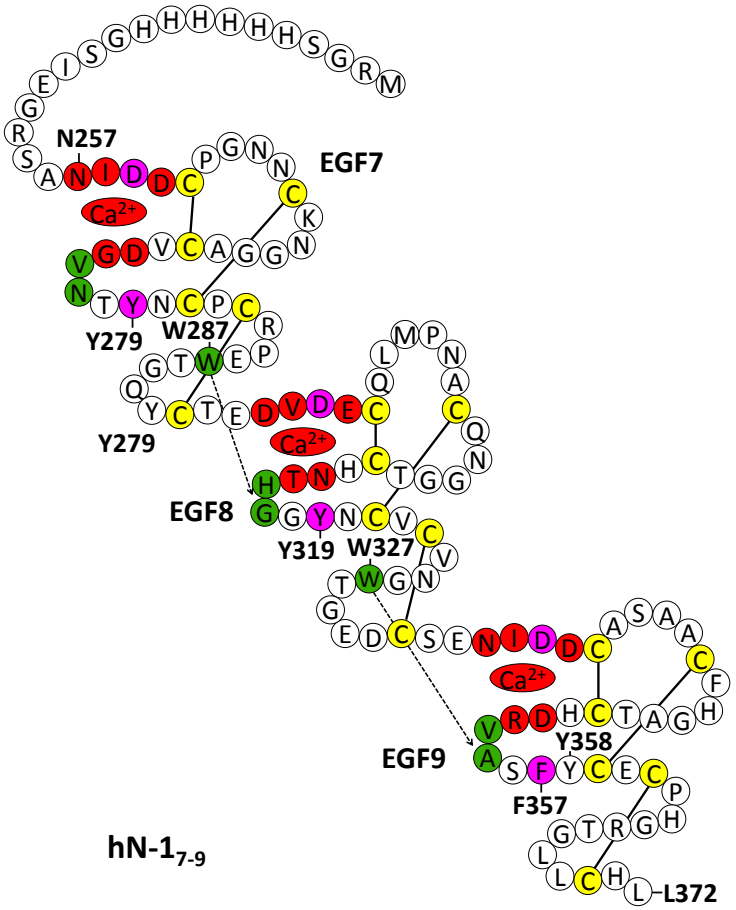


Figure 4.16 Schematic diagrams of hN-1_{7-9}. Aromatic markers are indicated, calcium and calcium binding ligands in red, packing residues in green, conserved residues for intradomain stability in purple, and cysteines in yellow. Because of a cryptic cleavage site for factor Xa in EGF9, the his-tag was not removed.

All three calcium binding sites are saturated at 3 mM Ca^{2+} (Figure 4.17). Y319, the consensus aromatic residue in EGF8, is saturated at an equimolar Ca^{2+} concentration, revealing high affinity in EGF8. Intermediate affinity is observed for EGF9, as peaks belonging to aromatic consensus residue F357 show an intermediate loss of intensity at 300 µM Ca^{2+}. Y279, the aromatic consensus residue in EGF7, shows shift changes with higher concentrations of Ca^{2+} for EGF7. However, saturation at 3 mM Ca^{2+} suggests a relatively high affinity for this N-terminal domain.

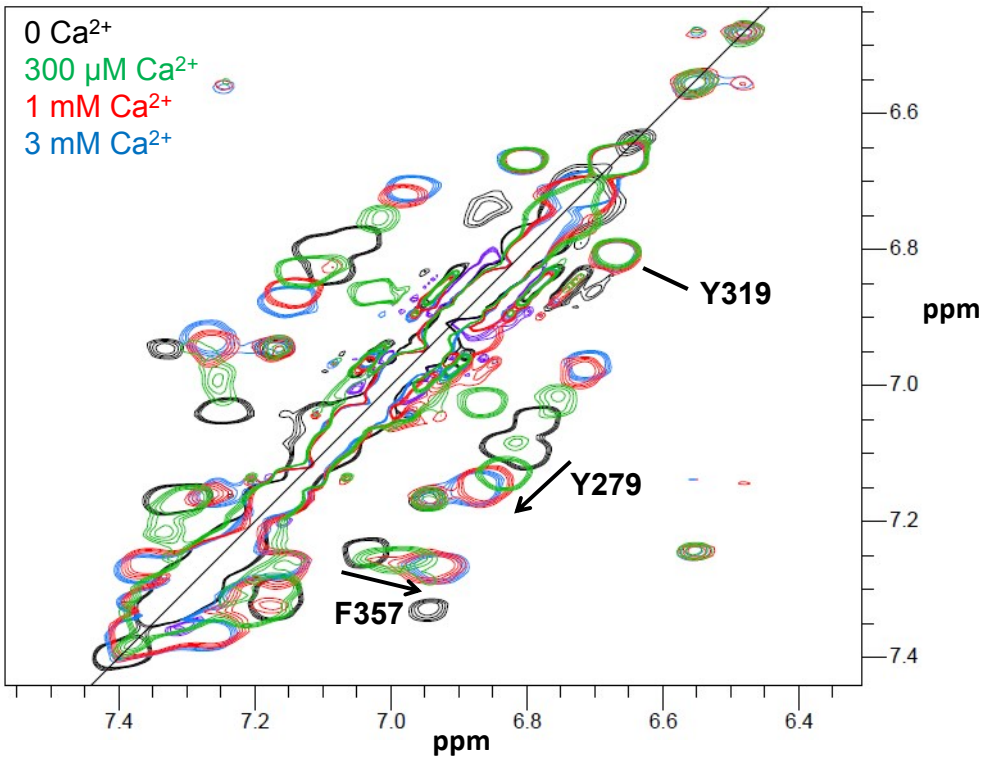


Figure 4.17 2D NOESY Ca^{2+} titration of 300 µM hN-1_{7-9}, at pH 7.5 and *I* = 0.15. Arrows indicate the direction of chemical shifts during the titration. Note that all 3 Ca^{2+} binding sites are saturated at 3 mM Ca^{2+}.

Ca^{2+} chelation experiments were performed to determine the K_d values for EGF8 and EGF9 (Figure 4.18). The dataset was fitted to one and two-site binding.

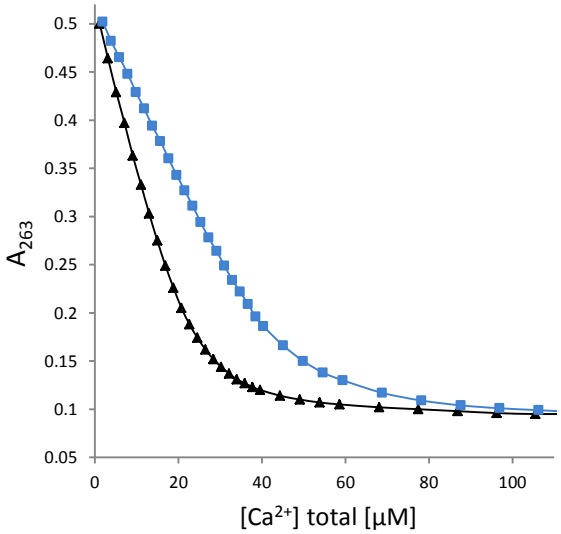


Figure 4.18 Ca^{2+} titration curve of 5,5-Br_2BAPTA and hN-1_{7-9} at pH 7.5 and *I* = 0.15. Ca^{2+} titration without protein (black triangles) shows the changes in absorbance upon Ca^{2+} addition. Addition of hN-1_{7-9} (blue squares) showed a stronger competition than other constructs (Figure 4.8 and Figure 4.13).

The error for two-site binding was statistically significant lower. The calculation of Ca^{2+} binding affinity coupled with the assignments based on NMR (Figure 4.17) revealed two high affinity sites with K_d values of 19 ± 3 µM for EGF9 and 1.1 ± 0.2 µM for EGF8 (n = 4, independent protein preparations).

Figure 4.19 summarizes Ca^{2+} affinities measured in the hN-1_{7-9} construct. The Ca^{2+} affinity for EGF7 was not measured because of the high error in free Ca^{2+} resulting from 3 cbEGF domains.

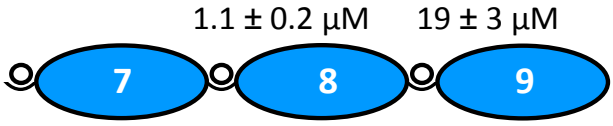


Figure 4.19 Ca^{2+} affinities in hN-1_{7-9}. Note that the affinity for EGF8 is around two orders of magnitude higher than in hN-1_{8-11}. The affinity for EGF9 at 19 ± 3 µM is within the error range of 26 ± 6 µM in EGF9 of hN-1_{8-11}.

4.6 Discussion

Figure 4.20 shows an overview of known Ca^{2+} dissociation constants in hN-1. In this study, K_d values for EGF5, EGF7, EGF8, EGF9 and EGF11 were measured. K_d values obtained for EGF11 in a native context, in chelation experiments with the chromophoric chelator 5,5-Br_2BAPTA were around 5 µM in all measured constructs (hN-1_{8-11}, hN-1_{10-14} and hN-1_{9-11}), demonstrating a reasonable precision for this method. A larger discrepancy was found for K_d values at the lower end of sensitivity of 5,5-Br_2BAPTA, such as 24 ± 4 µM for EGF7 in hN-1_{4-7} as compared to 16 ± 8 µM in hN-1_{5-7}. Similarly, the K_d for EGF9 was 19 ± 3 µM in hN-1_{7-9} and 26 ± 6 µM in hN-1_{8-11}. Due to the amount of spectrometer time needed to perform Ca^{2+} titrations by NMR, K_d measurements by 2D NOESY NMR spectroscopy were not repeated and the precision and accuracy of these values was not assessed. However, the individual titrations points were positioned well in the fitting curve for EGF5 in hN-1_{4-7} and hN-1_{5-7} (Figure 4.9 A and B) and for EGF9 in hN-1_{9-11} (Figure 4.14 A).

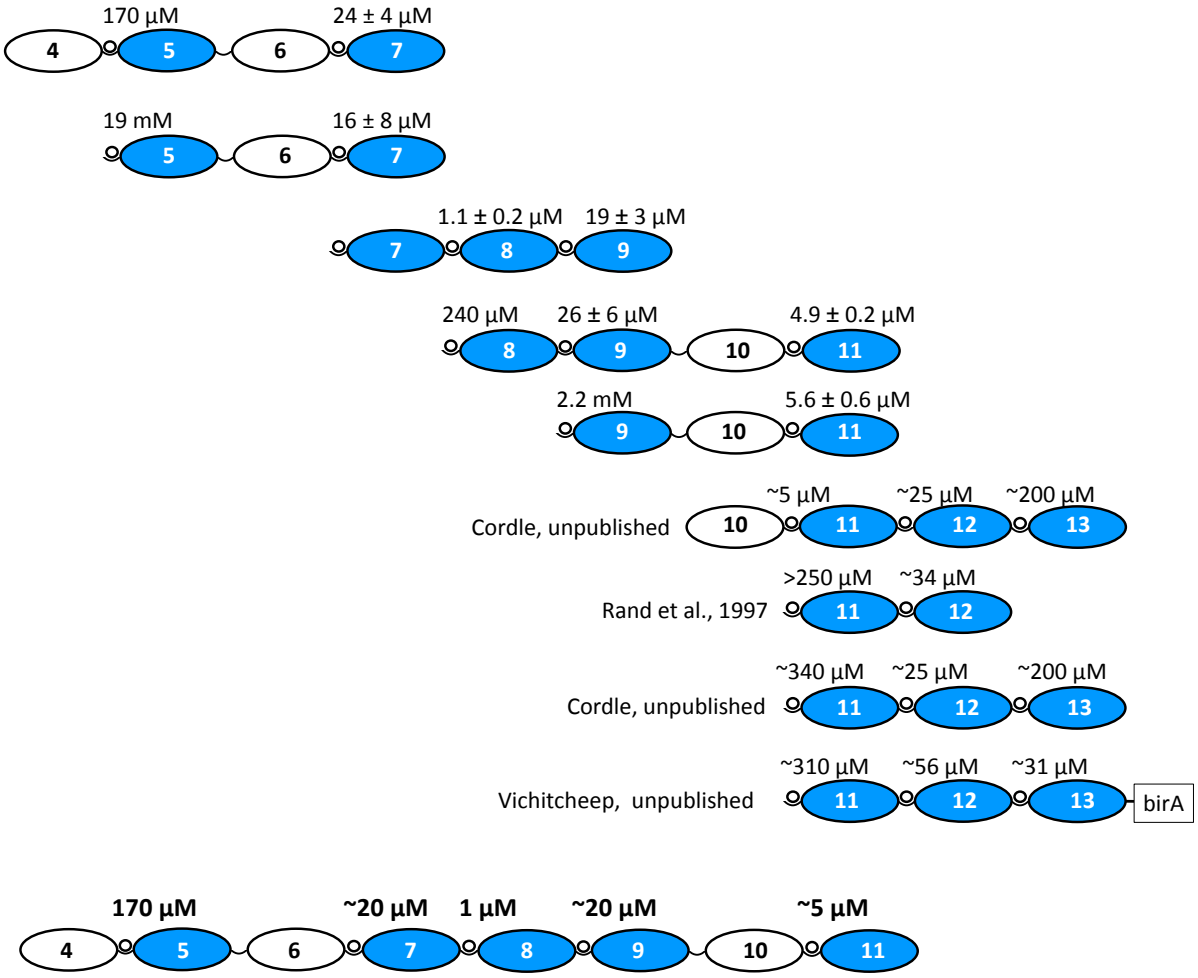


Figure 4.20 Overview of Ca^{2+} dissociation constants measured in different hN-1 constructs at pH 7.5 and *I* = 0.15. The measured dissociation constants are compared to unpublished data (Cordle, 2005, and Vichitcheep 2009) and published data (Rand et al., 1997). (Bottom) A summary of K_d values is plotted on the N-terminal region of Notch-1 (EGF4-13). K_d values for EGF13 ranged from 31 µM (Vichitcheep) with the C-terminal birA tag in hN-1_{11-13} to ~200 µM (Rand and Cordle) without the stabilising BirA tag.

On the other hand, the titration points during the fit for EGF8 of hN-1_{8-11} were less precise (Figure 4.14 B), most likely due to a relatively high affinity for an N-terminal domain and difficulties of calculating $[Ca^{2+}]_{free}$ in a domain construct with 3 cbEGF domains. Differences in K_d values obtained for EGF13 in hN-1_{11-13} by Vichitcheep (~31 µM) as compared to Cordle (~200 µM) are most likely due to the presence of a stabilising C-terminal tag (BirA tag) in the constructs used by Vichitcheep.

Adding a further cbEGF domain to the N-terminus of cbEGF12 in fibrillin-1 increased the affinity for Ca^{2+} by around two orders of magnitude (Smallridge et al., 1999). Here, the affinities of various Notch-1 constructs have been tested and the results confirm an increase of around two orders of magnitude for Ca^{2+} affinity in Notch-1 constructs in domains EGF5, EGF8, EGF9 and EGF11, when an N-terminal domain is added. This hundred fold increase in affinity suggests strong interactions of cbEGF domains with the preceding domains within the receptor. These results therefore agree well with the finding that Ca^{2+} binding rigidifies the fibrillin-1 cbEGF32-33 interface on a picosecond to microsecond timescale (Downing et al., 1996). Furthermore, they confirm that Ca^{2+} affinities are not affected by the length of the linker in between EGF domains (1 residue in class 1 cbEGFs as found in fibrillin-1, and two residues in class 2 EGF domains as found in Notch (Downing et al., 1996). The affinities obtained in this study add further structural information as the domain interactions in Notch might play an important role during ligand binding and receptor activation. If Ca^{2+} affinities *in vivo* are comparable to those measured here, most Ca^{2+} binding sites in the N-terminal region from EGF5 to EGF13 will be close to saturated with calcium at a physiological free calcium concentration of ~1 mM and pH 7.5 (Table 4.2).

Table 4.2 Saturation of different Ca^{2+} binding in sites in hN-1 at 1.0 mM Ca^{2+}_{free}.

domain	K_d (µM)	Saturation [%]
EGF-5	170	85.5
EGF-7	20	97.1
EGF-8	1	99.9
EGF-9	20	97.1
EGF-11	5	99.5
EGF-12	40	97.1
EGF-13	30	97.1

However, only around 85% of EGF5 Ca^{2+} binding sites in the receptor population will be saturated. If the Ca^{2+} binding in EGF5 is physiologically relevant, this relatively low Ca^{2+} affinity

may introduce heterogeneity within the Notch receptor population on the cell surface at a given time. This would make EGF5 more sensitive to small changes in free Ca^{2+} concentrations. Interestingly, a Ca^{2+} gradient has been linked to Notch signalling (Raya et al., 2004). Similarly, each Notch receptor molecule will have Ca^{2+} bound to EGF5 for 85% of the time at 1 mM free Ca^{2+} concentration. It is interesting to speculate whether or not the remaining 15% time, is physiologically relevant. Transient flexibility could occur in between EGF4-5 and to a lesser extent in the cbEGF domain pairs with higher affinity.

Interestingly, the lowest Ca^{2+} affinity measured in Notch so far is in EGF5 and this site is the very N-terminal cbEGF domain. EGF5 is further surrounded at the N-terminus by 4 non-calcium binding domains (EGF1-4) and followed by another non-calcium binding domain (EGF6, Figure 4.1) whereas EGF domains 7 to 23 are mostly Ca^{2+} binding. This clearly demonstrates that there is less Ca^{2+} binding at the N-terminus that potentially leads to a more flexible region. This flexibility at the N-terminus might allow the receptor to temporarily adopt a certain conformation allowing the formation of a supra molecular structure such as the Notch-ligand complex or an even higher order complex consisting of various ligand and receptor molecules (induced fit theory). Once the adopted conformation is held in place by additional forces, such as forces provided by the binding partners, Ca^{2+} may again bind to EGF5 and rigidify the interaction complex. This might be important in endocystosis where pH plays a role.

In conclusion, these results suggest rigid interfaces for the domain pairs EGF6-7, EGF7-8, EGF8-9, EGF10-11, EGF11-12 and EGF12-13 for human Notch-1 on the cell surface, but a potentially less rigid interface for EGF4-5. Further experiments are needed to understand the biological implication of the measured affinities. In addition, the effect of posttranslational modifications, such as sugars, β-hydroxylation has not been addressed in this study.

5 Assignments of chemical shifts

5.1 Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy uniquely allows the measurement of chemical properties for specific nuclei in molecules up to ~50 kDa, a feature due to the very distinct frequency of each nucleus. The assignment of each nucleus to a peak in the spectrum is thus the prerequisite for any NMR study.

The peak position is, by convention, described as the chemical shift in parts per million (ppm) as the difference between the observed frequency and the frequency of a standard, divided by this standard frequency; this removes the field strength dependence of the peak position. Figure 5.1 shows the 1D ^{1}H spectrum of hN-1_{4-7}. Peak positions correlate with the local magnetic field around each nucleus, which is the applied magnetic field B_0 minus the induced magnetic field B_i.

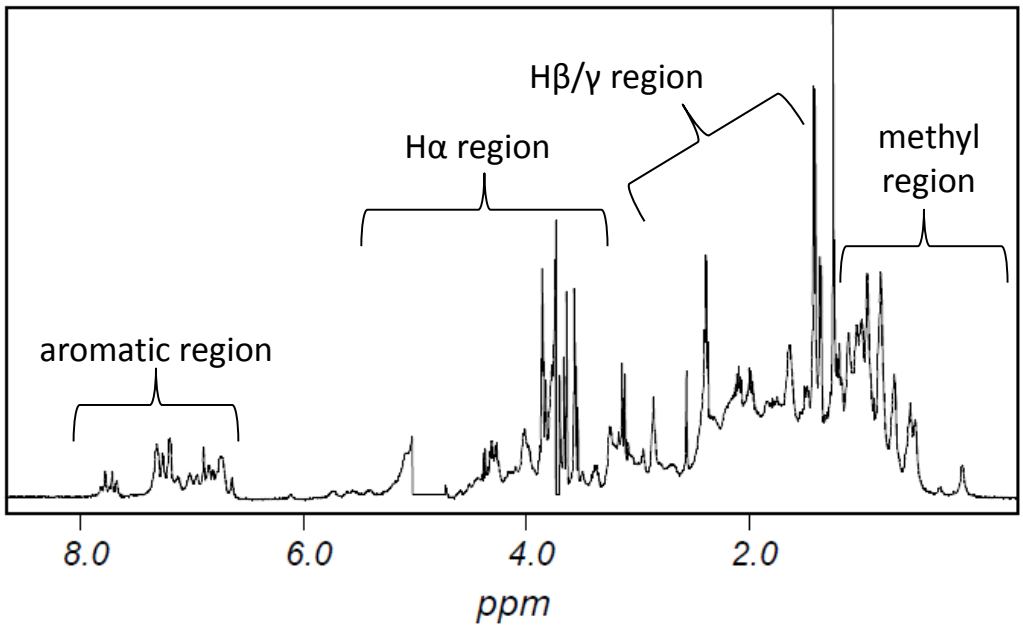


Figure 5.1 ^{1}H 1D NMR spectrum of hN-1_{4-7}. The different regions of the spectrum are indicated. Note that the large water peak at 4.75 ppm and Tris peak at 3.7 ppm were removed from the spectrum. The spectrum was recorded in 99.9% D_20, 5 mM Tris pH 8.3, 150 mM NaCl at 25 °C. Typical positions for peaks belonging to the aromatic region, the methyl region and the side chain region are indicated.

Nuclei with an extensive electron cloud, such as methyl groups, are more shielded from the applied magnetic field and their position in the spectrum is upfield around 0.5-2 ppm. Aromatic nuclei are less shielded from the magnetic field and their position is downfield at around 6-8 ppm. Assignment of the spectra of large molecules, such as proteins, is often challenging as the long overall rotational tumbling time results in fast relaxation and loss of signal intensity. Further difficulties arise from the large number of possible assignments for each resonance; proteins will contain a large number of residues of the same type. Thirdly, peaks are more likely to overlap as the size of the protein increases. Here, assignment strategies for proton, carbon and nitrogen resonances of the relatively large EGF domain constructs hN-$1_{4\text{-}7}$ (16.6 kDa), and hN-$1_{8\text{-}11}$ (18.7 kDa) are described. These constructs contain 4 EGF domains and hence are significantly larger than previously assigned double or triple EGF domain constructs (Hambleton et al., 2004, Malby et al., 2001). The assignment might also be more challenging because of the presence of sites of potential flexibility (EGF6 and EGF10, see figure 1.16). Whereas 1D ^{1}H and 2D ^{1}H-^{1}H NOESY experiments were used to characterise the proper fold of these constructs (chapter 3) and measure Ca^{2+} affinities (chapter 4), full assignments require triple resonance 3D NMR in which the peaks are further resolved using isotope labelling and a third dimension.

5.2. Results

5.2.1 HSQC analysis

2D ^{15}N-^{1}H HSQC spectroscopy is often the first NMR experiment for a newly expressed protein. It requires labelling with ^{15}N ammonium chloride and shows ^{1}H-^{15}N correlation peaks for all backbone amides and some side chains containing nitrogen (Asn, Gln, Trp). In some cases, Ca^{2+} HSQC titrations can be used to assign resonances to specific EGF domains. Here, Ca^{2+}

titrations on hN-1_{5-7} were performed to distinguish which peaks are affected by Ca^{2+} binding to EGF5 (K_d for Ca^{2+} = ~19 mM) and EGF7 (K_d for Ca^{2+} = ~16 µM) (Figure 5.2).

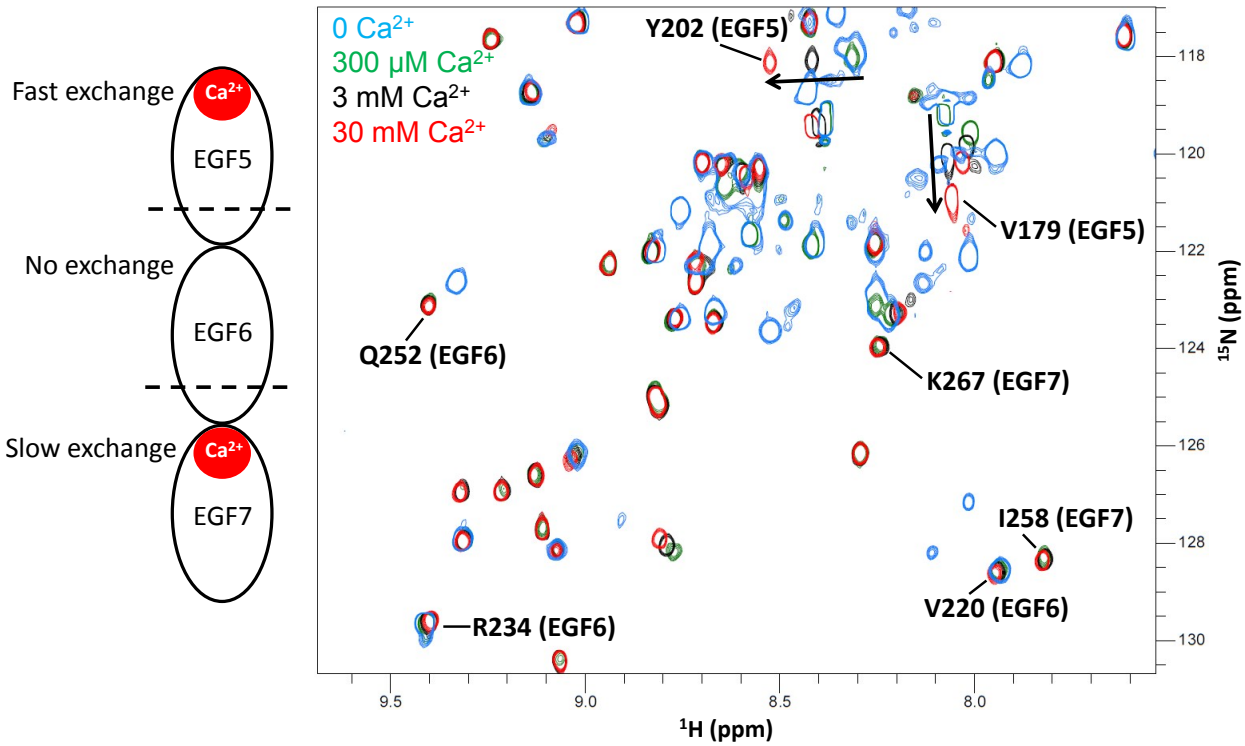


Figure 5.2 Ca^{2+} titration HSQC analysis of hN-1_{5-7}. The addition of an equimolar concentration of Ca^{2+} (green) to a sample with hN-1_{5-7} alone (blue) reveals residues belonging to EGF7 (I258 and K267) and the C-terminus of EGF6 (Q252) (slow exchange). Further addition of Ca^{2+} (3 mM in black and 30 mM in red) shows peaks that shift (fast exchange), revealing peaks belonging to EGF5 (V179 and Y202). Peaks not affected at all by Ca^{2+} belong to EGF6 (such as R234 and V220). Arrows indicate the direction of shift after the addition of Ca^{2+}. Spectra were recorded at pH 6.1, 25 °C in 5% D_2O and a protein concentration of 300 µM.

Peaks belonging to EGF7 (K267, I258) and the C-terminus of EGF6 (Q252) were easily identified as the residues that showed slow exchange behaviour (Figure 4.4, section 4.2.2) and appeared at the "bound" position after the addition of an equimolar amount of Ca^{2+} (300 µM in green, 0 Ca^{2+} in blue). Further addition of Ca^{2+} Figure 5.2 (3 mM in black, and 30 mM in red) showed fast exchange behaviour with shifting peaks for residues belonging to EGF5, examples are V179 and Y202. Peaks belonging to EGF6, such as R234 and V220, were not affected at all by Ca^{2+}.

5.2.2 Overlapping constructs

Assignment of the hN-1_{4-7} construct was facilitated using a dissection approach using shorter constructs. Figure 5.3 shows overlays of the hN-1_{4-7} construct (black) with hN-1_{4-5} (red, left hand panel), and hN-1_{5-7} (green, right hand panel).

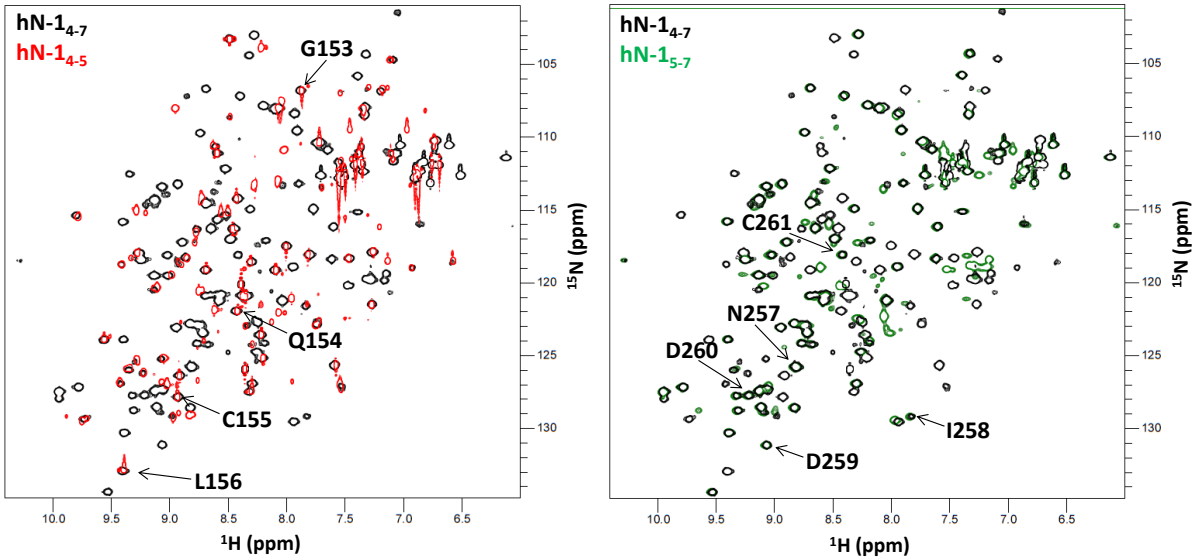


Figure 5.3 HSQC analysis of hN-1_{4-7} and smaller constructs. Shown are overlays of hN-1_{4-7} (black) and hN-1_{4-5} (red, left) and hN-1_{5-7} (green, right). Note that this dissection approach can be used to assign residues to certain regions in the construct. Residues that belong to EGF4 (left) or EGF7 (right) are indicated by arrows. Spectra were recorded at pH 6.1, 25 °C and 30 mM Ca^{2+} and protein concentrations of 1 mM.

Peaks belonging to EGF4 and the N-terminus of EGF5 superimpose in the left spectrum, while peaks belonging to the C-terminus of EGF5, EGF6 and EGF7 superimpose in the spectrum on the right. Peaks of residues belonging to either EGF4 (left) or EGF7 (right) are indicated by arrows. A third construct, hN-1_{4-6} was also made since certain residues of EGF5 were affected by the non-native context of EGF5 in hN-1_{5-7} and hN-1_{4-5}.

5.2.3 3D NOESY/TOCSY spectroscopy

^{15}N-^{1}H HSQC correlation peaks shown in Figure 5.3 were assigned using 3D gradient-enhanced ^{15}N-separated experiments, with the second ^{1}H dimension for protons close in space

(NOESY) or protons belonging to the same amino acid spin system (TOCSY). Figure 5.4 shows strip plots of residues N257-C261 at the beginning of EGF7 (NOESY black, TOCSY red).

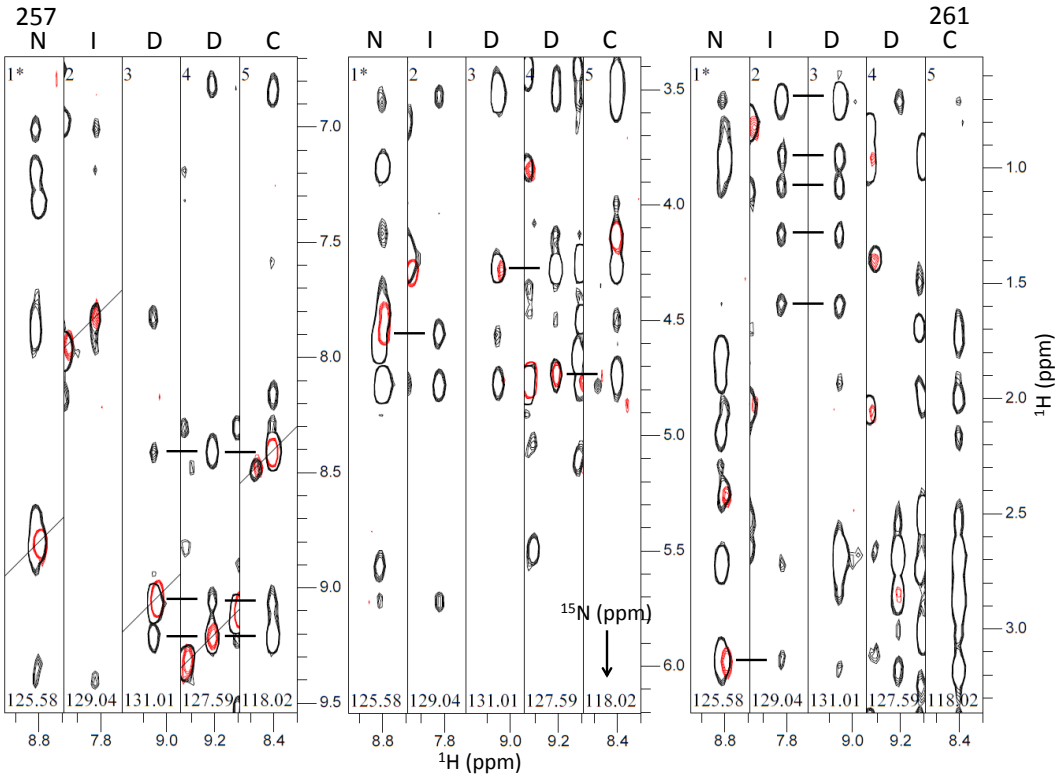


Figure 5.4 Overlay of 3D NOESY/TOCSY strips for N257-C261 in hN-1_{5-7}. ^{15}N-^{1}H TOCSY (red) indicates peaks belonging to the same residue while NOESY (black) peaks belong to protons close in space, inter-residue connections are indicated (black bar). Note that the nitrogen chemical shift is indicated at the bottom of each strip. The strips on the left show HN-HN NOEs, the middle set of strips show αH-NH NOE and TOCSY peaks, while the strips on the right show side chain connections. Spectra were recorded at pH 6.1, 25 °C, 30 mM Ca^{2+} and protein concentrations of 1 mM.

The plots shown for the hN-1_{5-7} construct are of significantly better quality than for the hN-1_{4-7} spectra (not shown). ^{1}H-^{15}N NOESY/TOCSY experiments were also performed for hN-1_{4-5} and hN-1_{4-6}. Connections usually observed between sequential residues involve the backbone amide (left plot), and the Hα (middle plot) Hβ and Hγ (right plot) of the preceding residue. However, these connections are also possible between non-sequential residues through space, as often seen for NH to NH backbone connections on opposing strands of a β-sheet. The presence of strong TOCSY signals can help to identify or narrow down amino acid type and potential position in the amino acid sequence; although in the constructs studied here, good quality 3D TOCSY spectra were

rare. For example, the side chain protons of I258 gave strong signals in the I258 and the D259 NOESY strips, but the absence of TOCSY signals initially impaired the assignment of I258. In conclusion, 3D ^{1}H-^{15}N NOESY/TOCSY spectroscopy in combination with HSQC analysis is an important first step in the assignment of a protein, but cannot provide full assignments in the constructs analysed here.

5.2.4 Triple resonance experiments

Labelling with more expensive ^{13}C allows the acquisition of triple resonance experiments. A combination of HNCA, CBCACONH, HNCO, HNCACO and HBHACONH pulse sequences was found most useful in confirming sequential residues suggested by 3D ^{1}H/^{15}N NOESY/TOCSY experiments and to obtain Cα/Cβ/Hα/Hβ and backbone carbonyl chemical shifts (Figure 5.5).

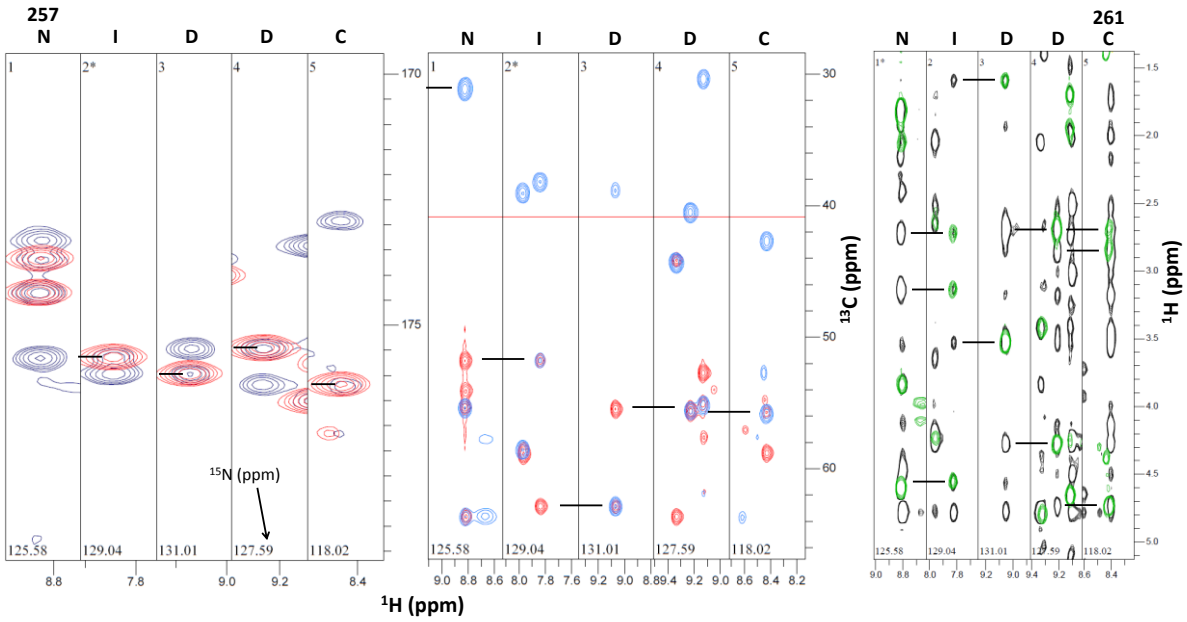


Figure 5.5 Overlay of 3D HNCA and CBCACONH (middle), HNCO and HNCACO (left), and HBHACONH and NOESY (right) in hN-1$_{5-7}$. The HNCA (red in the middle) with its good sensitivity confirmed sequential peaks suggested by the ^{1}H/^{15}N 3D NOESY. CBCACONH (blue) experiments provided the Cα/Cβ shifts of the preceding residue. Similarly, the overlay of the HNCO (red) and HNCACO (blue, left) shows the connection of carbonyl peaks for sequential residues. Side chain proton information was obtained for residues i-1 with HBHACONH (green) shown overlayed on the NOESY (black on the right). Multiple experiments confirm sequential residues, which is especially helpful with weak intensity or overlapping peaks. Spectra were recorded at pH 6.1, 25 °C, 5% D_2O, 30 mM Ca^{2+} and a protein concentration of 1 mM.

HNCA was the most sensitive experiment to correlate sequential residues, with peaks connecting the backbone NH to its own Cα and the Cα of the preceding residue (Figure 5.5 middle in red). Note the connections between N257, I258 and D259 not seen clearly in the previous section (3D ^{1}H-^{15}N NOESY/TOCSY spectroscopy). The Cα chemical shift was used to narrow down amino acid type, as in the case of the downfield shifted Cα of I258 (~63 ppm), characteristic of amino acids like I, P, V, S, T. Similarly, glycines have a very upfield shifted Cα around 40 ppm.

The CBCACONH (Figure 5.5 middle) was sensitive too and indicated the Cα/Cβ of residue i-1, with the Cβ especially helpful for residues like alanine with a very upfield shifted Cβ of 15 ppm. Similarly, sequential peaks can be confirmed using a combination of HNCO/HNCACO (Figure 5.5 left) that connects the backbone NH to the backbone carbonyl and that of the preceding residue (HNCO in red) and to its own carbonyl and that of the preceding residue (HNCACO, in blue).

Another experiment used successfully was the HBHACONH that provides the proton chemical shifts of Hα and Hβ of the preceding residue for each NH (right, green, overlayed with NOESY in black, Figure 5.5). The chemical shifts obtained for Cα/Cβ/Hα/Hβ from CBCACONH/HBHACONH experiments helped in assigning residues to a specific amino acid type.

Further side chain information was obtained from ^{1}H-^{13}C HSQC and NOESY/HCCH-TOCSY experiments. Figure 5.6 shows strips for the full side chain spin system of I258.

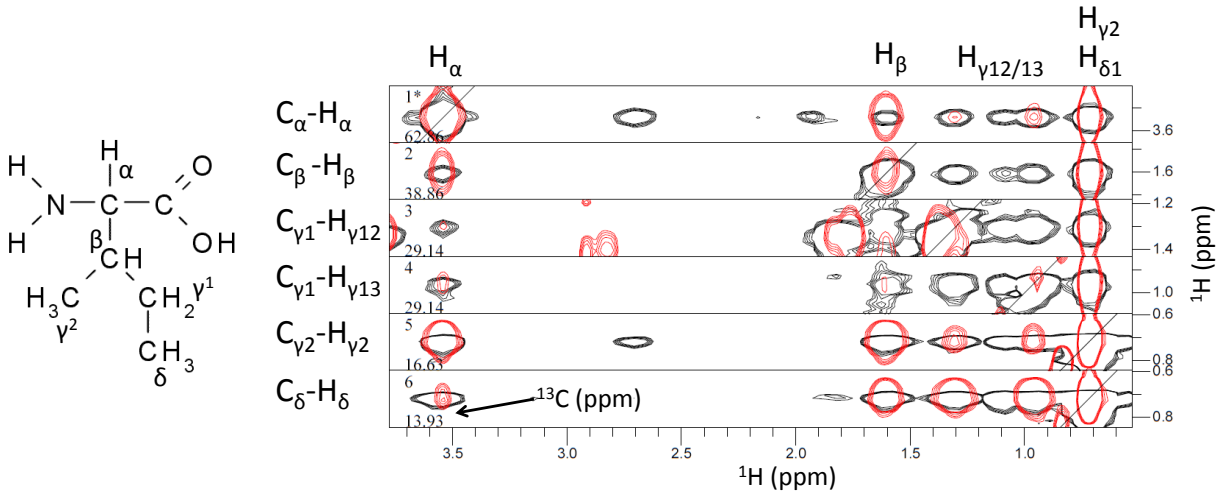


Figure 5.6 Overlay of 3D ^{13}C NOESY (black) and HCCH-TOCSY (red) strips for I258 in hN-1$_{5-7}$. Whereas NOESY peaks indicate resonances of protons in close proximity in space, the HCCH-TOCSY confirmed that these protons belong to the same spin system. The structure of isoleucine is indicated on the left. The ^{13}C chemical shift for each strip is indicated at the left of the strip. Spectra were recorded in 99.9% D_20, pH 6.1, 25 °C, 30 mM Ca^{2+} and a protein concentration of 1 mM.

This strategy, using a combination of various double- and triple-resonance experiments, was also used to obtain backbone and side chain assignments for hN-1$_{4-7}$.

5.2.5 Further optimisation and the hN-1$_{8-11}$ construct

Obtaining full assignments with the strategy explained in 5.2.4 was straightforward in the presence of strong signals for hN-1$_{4-7}$, when a dissection approach using smaller constructs (hN-1$_{4-5}$, hN-1$_{5-7}$ and hN-1$_{4-6}$) was adopted. The assignment of hN-1$_{9-11}$ under saturating Ca^{2+} concentrations (red in Figure 5.7), indicated that resonances of an entire EGF domain were missing, since the number of peaks in the spectrum corresponded to 2 EGF domains only.

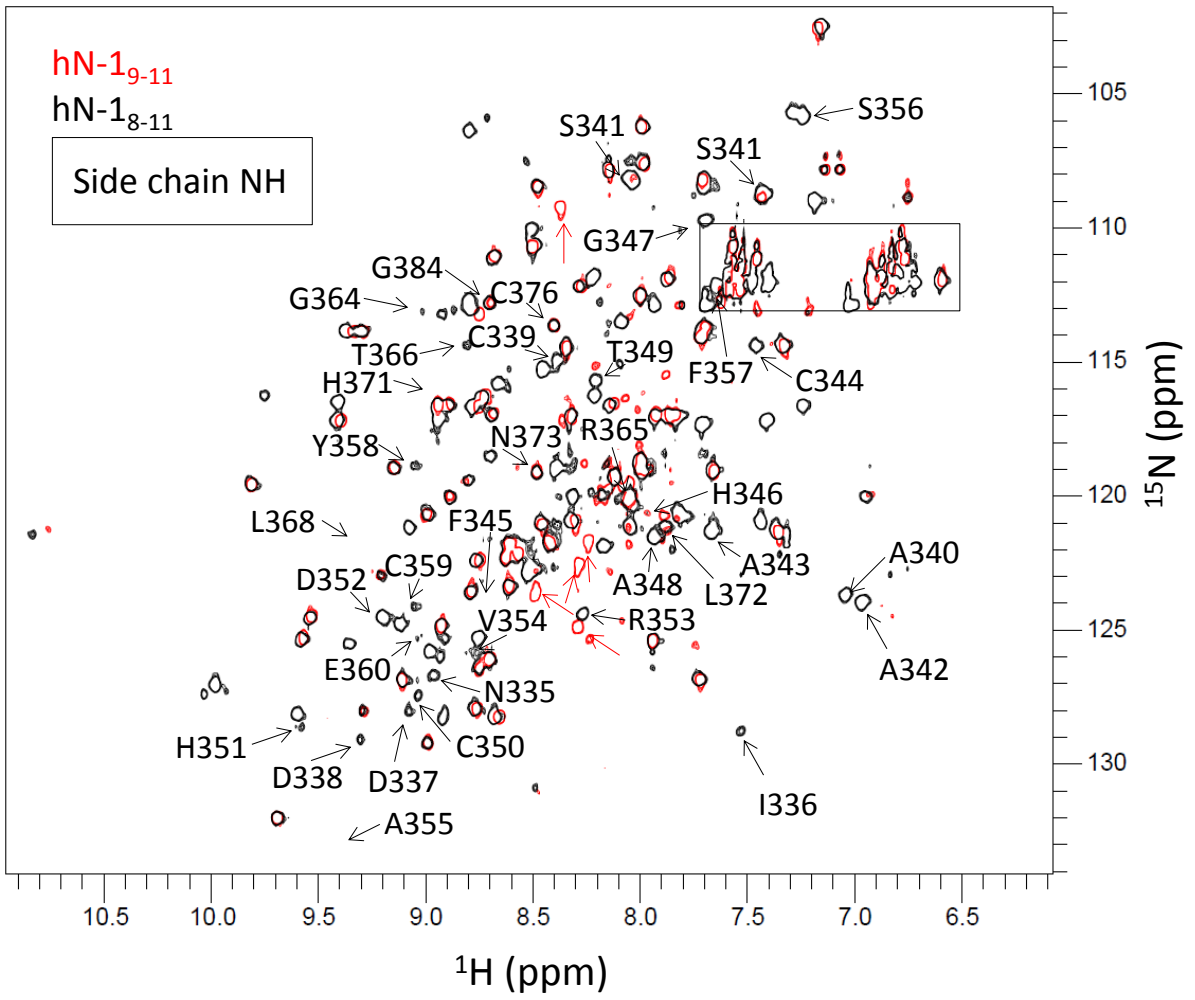


Figure 5.7 ^{15}N-^{1}H HSQC analysis of hN-1$_{9-11}$ (red) and hN-1$_{8-11}$ (black). Only peaks belonging to EGF9 are labelled; these were present in hN-1$_{8-11}$, but not in hN-1$_{9-11}$. This is likely due to the stabilising effect when EGF9 is placed in a native context with a preceding EGF domain (EGF8). Intense peaks observed for hN-1$_{9-11}$ but not hN-1$_{8-11}$ correspond to residues belonging to the His-tag of the hN-1$_{9-11}$ construct and are indicated with a red arrow (chapter 3). Spectra were recorded at pH 6.1, 25 °C and 30 mM Ca^{2+}.

The missing peaks belonged to EGF9 as the overlay with a spectrum of hN-1$_{8-11}$ (black) and subsequent assignment showed. The K_d of EGF9 for Ca^{2+}, when it is an N-terminal domain, was around 2.2 mM and hence had significantly weaker affinity than N-terminal EGF8 (K_d = 240 µM) or EGF11 (K_d = 300 µM) (chapter 4), a fact that could contribute to the lower stability of EGF9 in the hN-1$_{9-11}$ construct. Furthermore, peaks of EGF9 were broad in a C-terminal context, as observed for the hN-1$_{7-9}$ construct (not shown). In addition, unlike in the hN-1$_{11-13}$ construct

(Hambleton et al., 2004, Cordle et al., 2008a), the addition of the biotin ligase (BirA) tag at the C-terminus had no significant stabilising effect suggesting that this is an inherent property of EGF9.

As shown in the previous section, the prerequisite for full assignments is the presence of peaks with a reasonable intensity. At low protein concentrations, where protein is monomeric, peak intensities double as the concentration of protein is doubled. At high concentration, some proteins tend to aggregate or precipitate with a decrease in NMR signal intensity. Therefore, testing the optimum protein concentration is crucial. Here, an analysis for the optimisation of sample conditions for the hN-$1_{8\text{-}11}$ construct (30 mM Ca^{2+}, pH 6.1) was performed (Figure 5.8). Intensities of the HSQC peaks increased linearly up to a concentration of around 500 µM (not shown). A further increase of around 50% is seen after doubling the protein concentration to 1 mM protein. At 1.5 mM, the peak intensities are similar to 1 mM, and at 2 mM the signal intensities are decreasing. At 3 mM, many peaks continue to show a reduced intensity (Figure 5.8). Interestingly, a few peaks do become more intense at 3mM. These peaks correspond to residues with intrinsic flexibility as demonstrated by a low heteronuclear NOE value (later chapter). Table 5.1 shows representative examples of peak intensities. A420, a residue in the flexible loop between C1 and C2 of EGF11 shows an increase in peak intensity as does D295, a very flexible, N-terminal residue.

Table 5.1 Intensities of ^{15}N-^{1}H correlation peaks at 1 mM and 3 mM protein concentration in the hN-$1_{8\text{-}11}$ construct. Residues which are flexible on a fast timescale, such as A420 and D295 show increased intensity at 3 mM. Residues that are more rigid like V392 and G309 decrease their intensity. Intensity values are unitless.

	1 mM	3 mM	hN-NOE
D295	$3.5*10^7$	$5.3*10^7$	-0.155
G309	$4.7*10^6$	$2.0*10^6$	0.543
V392	$5.8*10^6$	$1.8*10^6$	0.739
A420	$4.9*10^7$	$6.6*10^7$	-0.042

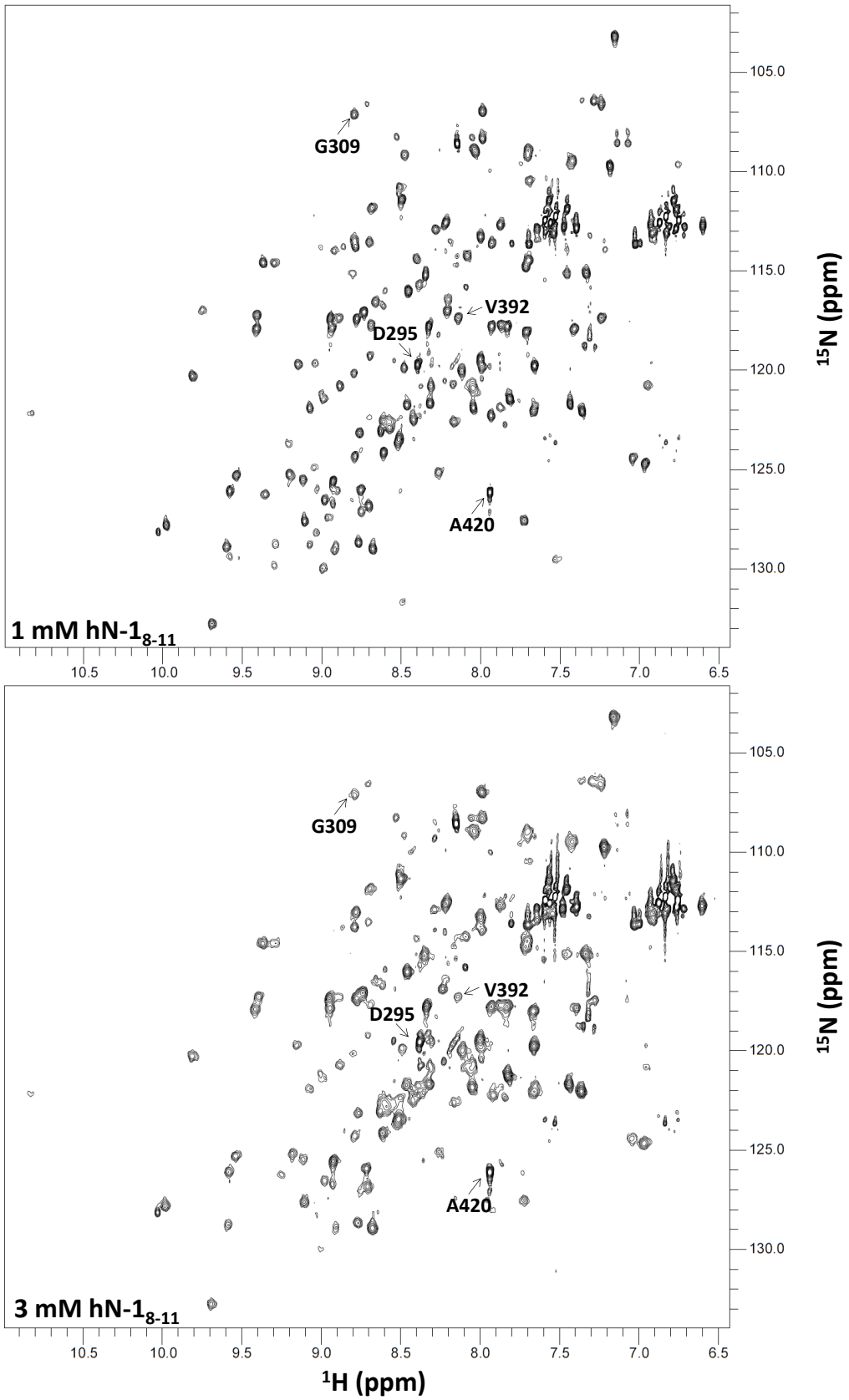


Figure 5.8 HSQC analysis of 1 mM (top) and 3 mM (bottom) hN-1$_{8-11}$. As for G309 and V392, most peaks are much better resolved at 1 mM than at 3 mM, demonstrating the need for aggregation tests before a full series of 3D spectra can be recorded. Resonances for residues with intrinsic flexibility, such as D295 and A420, do however increase in intensity despite aggregation. Spectra were recorded at pH 6.1, 25 °C and 30 mM Ca^{2+}.

This effect is due to the fact that the T_2 relaxation of very flexible residues is less affected by protein aggregation. On the other hand, residues G309 and V392 clearly lost intensity at the 3 mM protein concentration. Aggregation was not overcome by the addition of 100 mM arginine/glutamic acid as described previously (Golovanov et al., 2004), or the addition of 150 mM NaCl. No effect was seen after the increase in pH from 6.1 to pH 7.0.

Similar HSQC analysis of peak intensities was carried out for the hN-$1_{4\text{-}7}$ and hN-$1_{11\text{-}13}$ constructs, both of which gave an optimal spectrum at around 1 mM protein concentration. Other variables to adjust are the temperature and the length of an experiment. Usually, 3D experiments were recorded for 3 days. Since the signal to noise ratio (S/N) increases with the square root of the scans taken, the S/N could be doubled by using 4 times the number of scans, which would take 12 days of magnet time. The effect of the temperature and magnetic field strength was also tested. The best spectra were obtained using the 500 MHz Bruker spectrometer, most likely because exchange-based line broadening is less severe at low magnetic field and the availability of a cryoprobe on this spectrometer. A decrease from 25 °C to 10 °C decreased peak intensities since the overall correlation time increased and peaks broadened due to exchange contributions did not improve. Due to potential sample instability, higher temperatures were not tested. In conclusion, NMR experiments were performed at a protein concentration of 1 mM, at pH 6.1 and 5% D_2O. Ca^{2+} concentrations were 80 mM for hN-$1_{5\text{-}7}$ due to its low affinity (chapter 4) and 30 mM for the other constructs.

5.3 Conclusion and extent of assignment

Figure 5.9 shows an overview of the applied assignment strategy. Constructs consisting of 2 or 3 domains only, show relative strong intensities. However, their N- or C-terminal domains might not be stable because of low Ca^{2+} affinity such as in the N-terminal domain of hN-$1_{5\text{-}7}$ or because they are not in a native context, such as EGF9 in hN-$1_{9\text{-}11}$. Therefore, the larger

constructs hN-1_{4-7} and hN-1_{8-11} were generated and complemented with assignments obtained from the shorter constructs. The optimisation of the sample conditions, such as protein concentration was found to be an important step for good quality spectra.

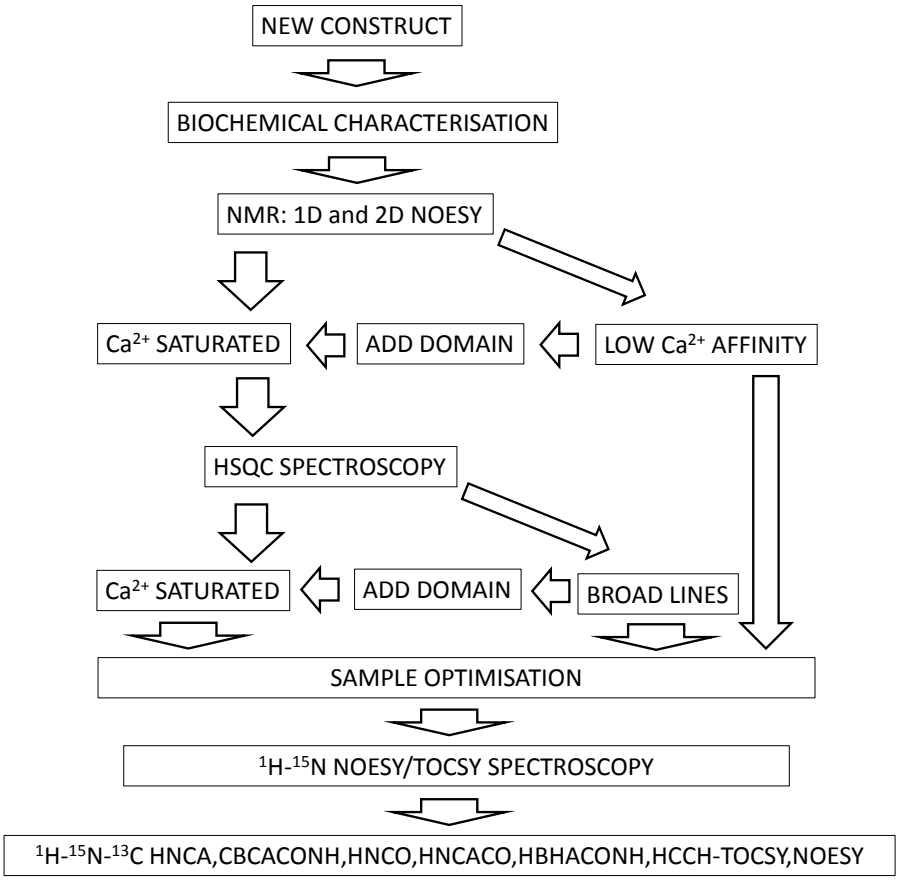


Figure 5.9 Assignment strategy used for different constructs in hN-1_{4-11}. Note that shorter constructs can have stronger intensities and are therefore helpful to complement assignments of larger constructs.

This strategy was applied to constructs hN-1_{4-7} and hN-1_{8-11}. Figure 5.10 shows the HSQC spectra of hN-1_{4-7}. All backbone NHs were assigned, apart from N180, E181, C182. Further assignments were 93.7% of Hα, 89.8% of Hβ, 38% of Hγ, 35.6% of Hδ and 29.3% of Hε, 91.6% of ^{13}C', 96.1% of ^{13}Cα, 92.6% of ^{13}Cβ, 30.6% of ^{13}Cγ, 18.2% of ^{13}Cδ. Similarly, Figure 5.11 shows the HSQC spectra of hN-1_{8-11} for which H316 and L369 were the only backbone NHs unassigned. Further assignments were 92.6% of Hα, 89% of Hβ, 62.1% of Hγ, 48.9% of Hδ and 32.4% of Hε, 96% of ^{13}C', 100% of ^{13}Cα, 95.7% of ^{13}Cβ, 42.5% of ^{13}Cγ, 40.3% of ^{13}Cδ and 15.8% of ^{13}Cε.

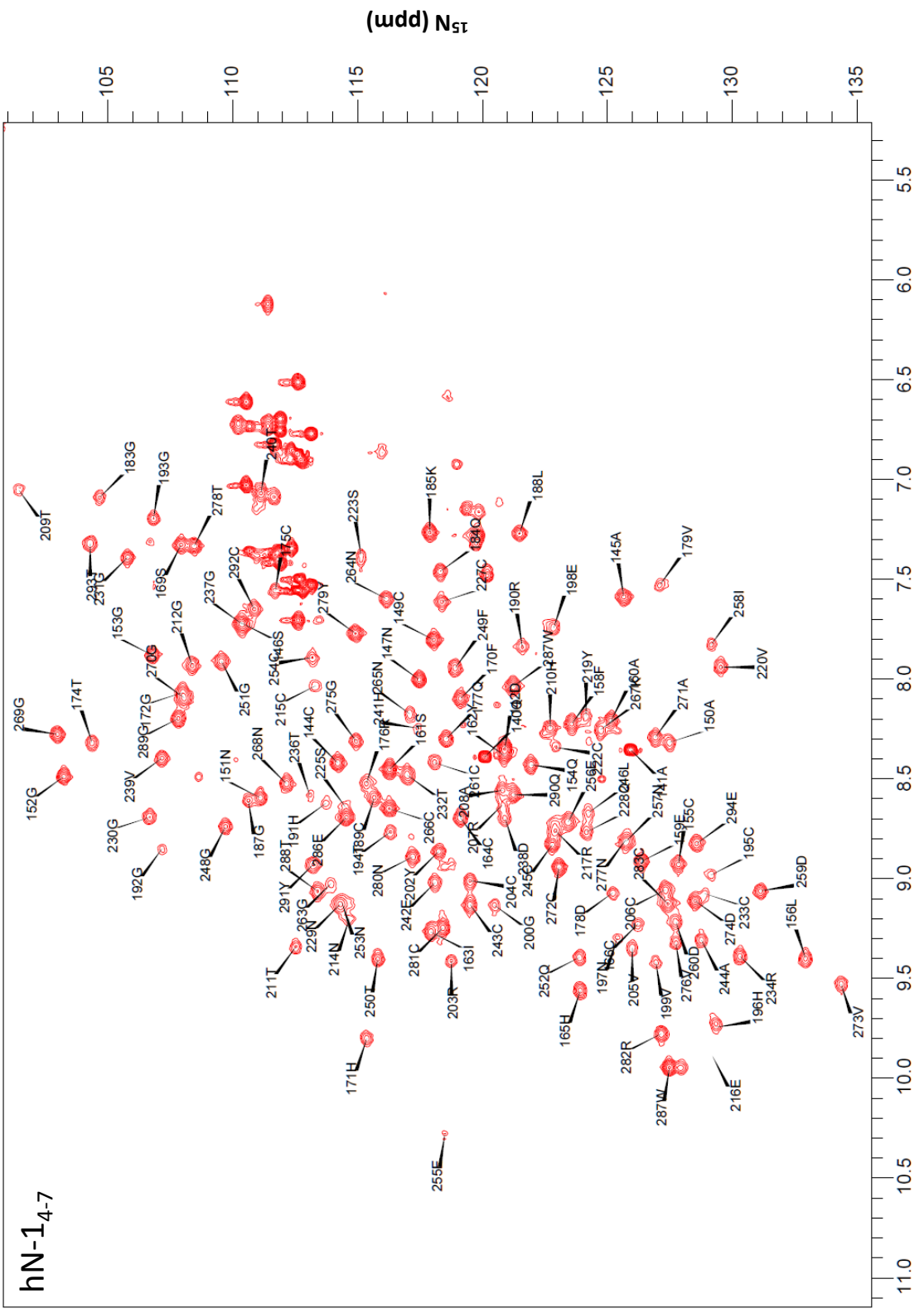


Figure 5.10 HSQC of hN-1$_{4-7}$, collected at 500 MHz (Bruker), 30 mM Ca^{2+}, pH 6.1, 1 mM protein concentration and 25 °C.

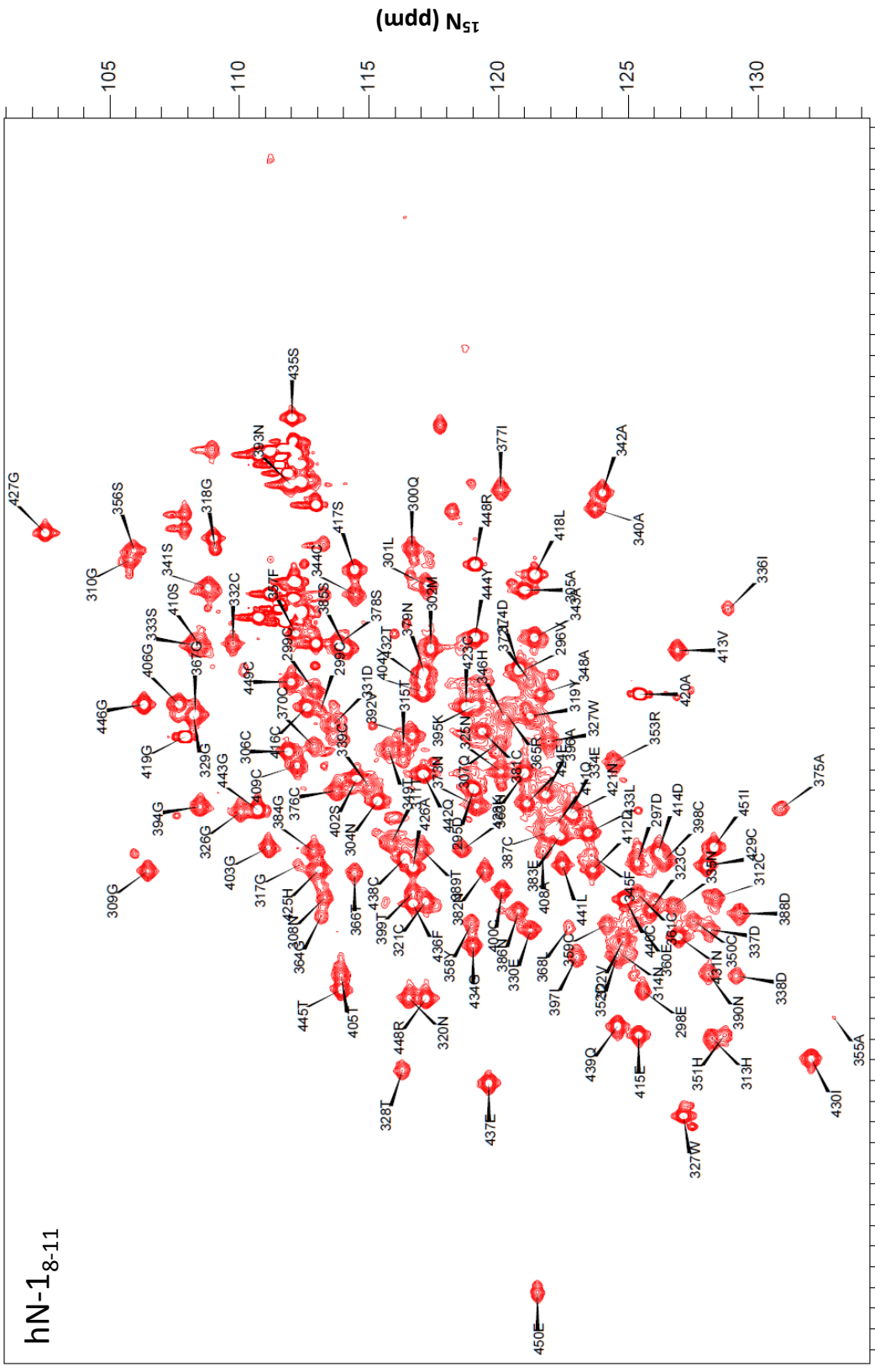


Figure 5.11 HSQC of hN-1$_{8-11}$, collected at 500 MHz (Bruker), 30 mM Ca^{2+}, pH 6.1, 1 mM protein concentration, and 25 °C.

Figure 5.12 shows the HSQC collected for hN-1$_{7-9}$, a construct that bridges hN-1$_{4-7}$ and hN-1$_{8-11}$. All backbone NHs were assigned, apart from V324 and N325. Side chains were not

assigned in this construct since sufficient side chain assignments were available from hN-1_{4-7} and hN-1_{8-11}.

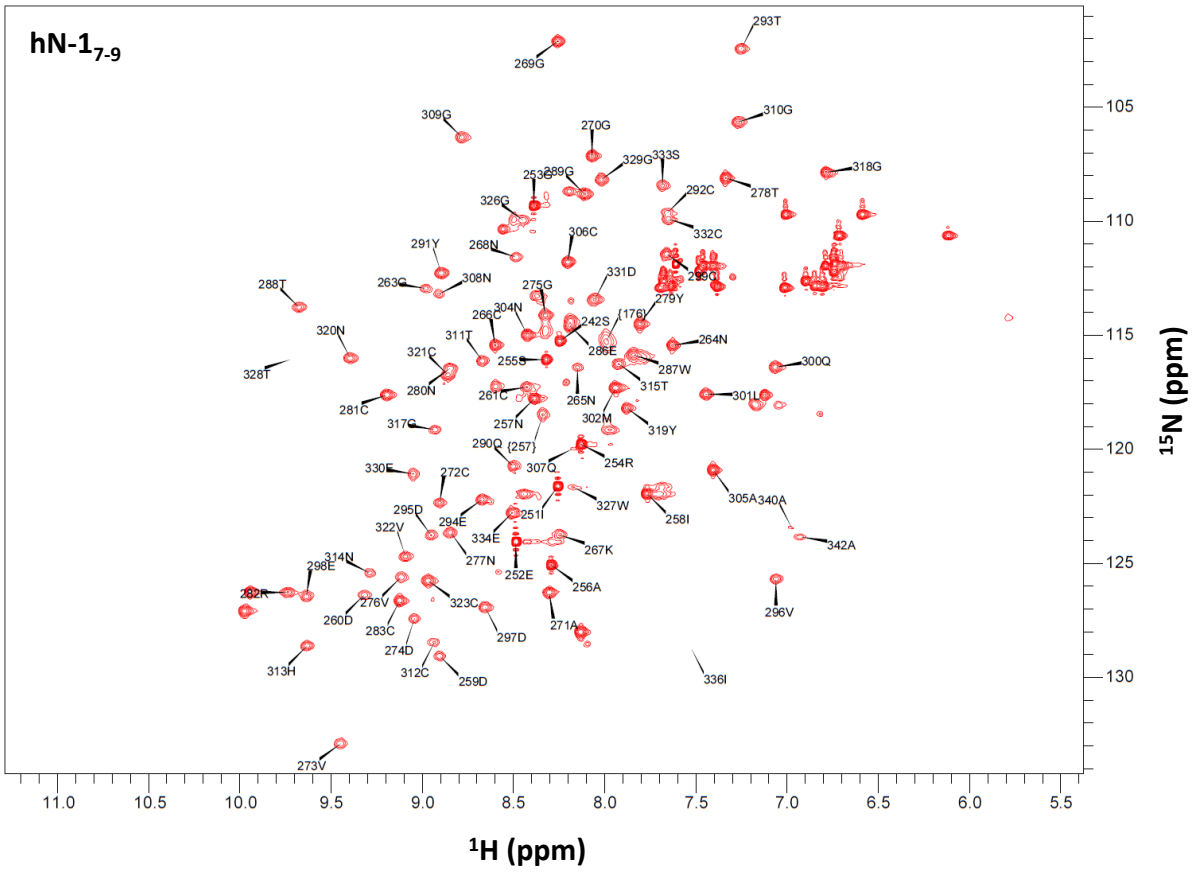


Figure 5.12 HSQC of hN-1_{7-9}, collected at 500 Mhz (Bruker), 30 mM Ca^{2+}, pH 6.1, 25 °C and a protein concentration of 1 mM.

Assignments are presented in the appendix and represent the starting point for subsequent NMR analysis.

In chapter 6, NOEs will be used to confirm interdomain contacts. The program TALOS will be used to calculate phi/psi backbone torsion angles (chapter 7). In chapter 8, the domain orientation will be analysed using residual dipolar couplings (RDC).

6 Structural analysis

6.1 Introduction

Despite the importance of the Notch signalling pathway, little is known about the cell-surface organisation of the Notch receptor and how the formation of the ligand-binding complex leads to endocytosis and receptor activation.

In the inactive state of Notch, the γ-secretase cleavage site (S2) is shielded by three LNR repeats in a cap-like arrangement (Figure 1.7). Although ligand binding initiates the opening of this cleavage site, soluble ligand was not able to activate Notch (Sun and Artavanis-Tsakonas, 1997). Similarly, Notch needs to be internalised by the opposing, ligand-expressing cell before activation can occur (Parks et al., 2000). These data were interpreted in a mechanotransduction model in which force on Notch generated through the endocytosis process detaches the LNR repeats from the S2 cleavage site.

In a deletion mapping approach, EGF11-12 were shown to be sufficient and required to cause ligand binding in cell aggregation assays, but they were not sufficient for Notch signalling (Rebay et al., 1993, Rebay et al., 1991). Binding measurements indicated an increase of affinity of the Notch-ligand complex when EGF domains were added outside the EGF11-12 region (Cordle et al., 2008b, Shimizu et al., 1999, Shimizu et al., 2000). These findings together with the mechanotransduction model suggested that the Notch-ligand complex requires a certain affinity to withstand the pulling force of the ligand-binding cell. These data therefore suggest a crucial role for EGF domains outside the core ligand-binding region EGF11-12. In this chapter, structural data for the extracellular domain of Notch-1 (Figure 6.1) is analysed. The structure of the core ligand-binding region (EGF11-12) has been solved in the hN-$1_{11\text{-}13}$ construct. This structure revealed a linear, rod-like conformation, stabilised by Ca^{2+} (Cordle et al., 2008a, Hambleton et al., 2004). A linear domain orientation has also been found in all other tandem cbEGF domain constructs studied to date (Smallridge et al., 2003, Abbott et al., 2007, Downing et al., 1996).

Since many cbEGF domains show a high degree of sequence conservation (Table 1.1), a linear-organisation for the Notch receptor was proposed. However, three EGF domains within the central region of the EGF domain rich region EGF6, EGF10, and EGF22 do not bind Ca^{2+} and were therefore predicted to confer flexibility to the receptor (Hambleton et al., 2004). In this chapter, recent X-ray crystallographic structures, solved in collaboration by Professor Susan Lea at the Sir William Dunn School of Pathology, Oxford, are used to investigate the structure of interfaces formed by EGF and cbEGF domains to identify whether or not these regions are flexible or rigid.

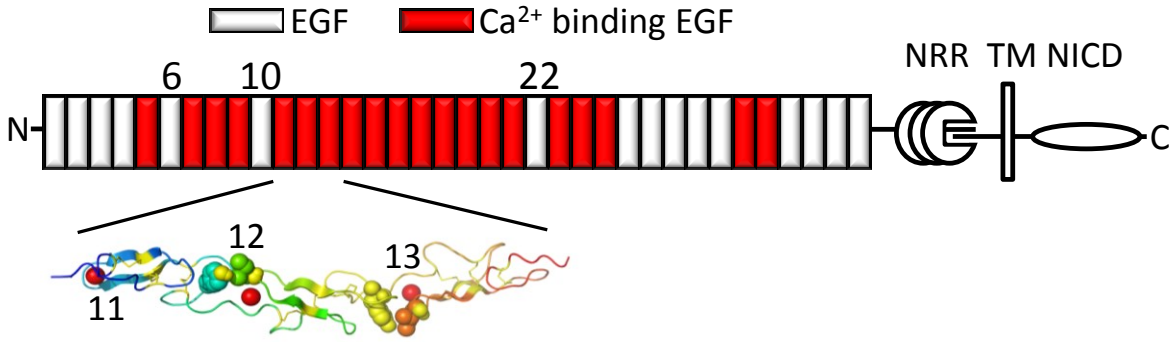


Figure 6.1 Extracellular domain of human Notch-1. Most EGF domains are Ca^{2+} binding (red). The X-ray structure of hN-1_{11-13} reveals a rod-like, elongated shape, stabilised by Ca^{2+} (red); packing residues indicated by spheres.

This will form a platform to understand data obtained in ligand binding and receptor activation studies. For example, adding EGF10 to the hN-1_{11-14} construct was shown to reduce binding to the Delta-like 1 ligand (Cordle et al., 2008b), the structural reasons for this remains unclear. Another study showed competition between Delta and the Abruptex region of Notch (EGF24-29) (Pei and Baker, 2008).

6.2 Structural analysis of hN-1_{11-13}

6.2.1 Ca^{2+} binding in hN-1_{11-13}

The structural conservation of the interfaces formed by EGF11-12 and EGF12-13 in the hN-1_{11-13} construct was analysed, using the X-ray coordinates (PDB ID: 2VJ3) and the Pymol align

function; the overlay of EGF12 on EGF13 shows the high degree of conservation in EGF domains observed earlier (Figure 6.2) (Bork et al., 1996). Although the major β-sheet and disulphide bonds in EGF12 and EGF13 overlay very well (top panel), a small difference is observed in the position of the Ca^{2+} ion and of the turn at the end of the β-sheet (arrow). Similarly, the position of EGF11 relative to EGF12 is slightly offset when EGF12 and EGF13 are overlayed (arrow).

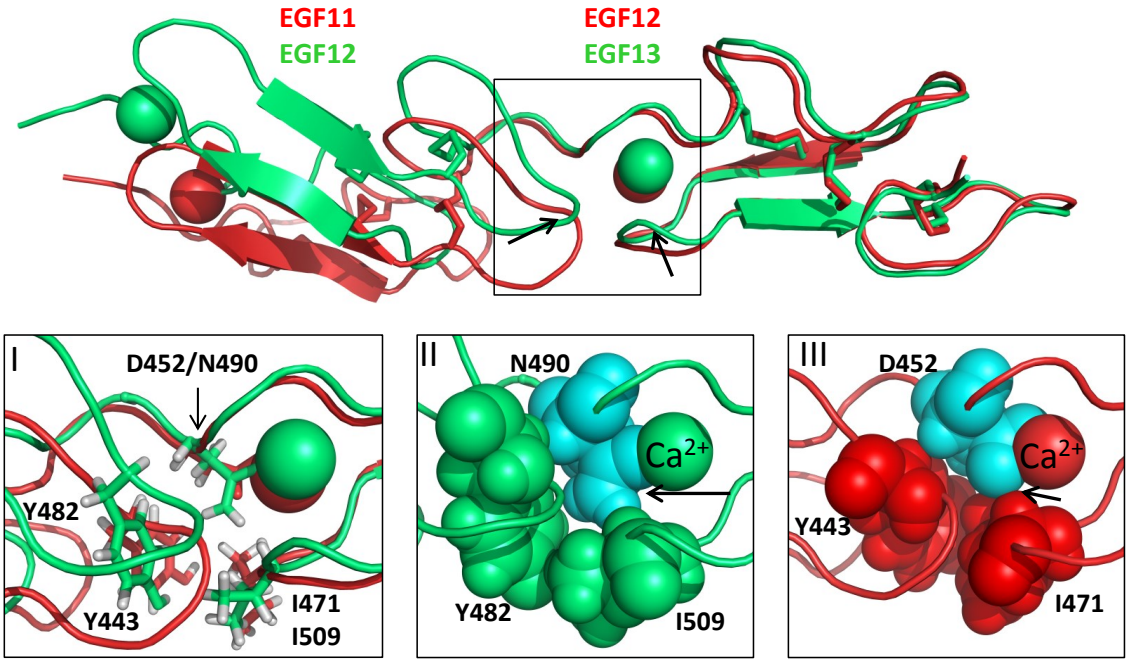


Figure 6.2 Overlay of domain pairs EGF11-12 (red) on EGF12-13 (green). The overlay (top panel) shows the structural conservation of EGF12 and EGF13, with little difference in the position of the Ca^{2+} ion (red in EGF12, green in EGF13) and the β-turn (arrows). Box I shows the conserved aromatic packing residues (Y443/Y482), the corresponding hydrophobic packing residues at the turn of the β-sheet in the following domain (I471/I509) and the side chains of the first Ca^{2+} ligand (D452/N490). Note that the side chain of N490/D452 is a Ca^{2+} ligand and therefore less flexible. In EGF13, the amide group of N490 (Box II, cyan and arrow), when compared to the carboxylate group of D452 (EGF12, Box III, cyan, arrow) points away from the Ca^{2+} ion and leads to an offset of Y482 and I509. This causes a slightly different interface, which has been referred to as "gently curved" (Cordle et al., 2008a).

Interestingly, both domain pairs have the same residue types at the interface (Box I) – a tyrosine (Y443 in EGF11, Y482 in EGF12) packs with an isoleucine (I471 in EGF12 and I509 in EGF13). However, another residue in close contact has a fixed orientation because of its function as a Ca^{2+} ligand. This is the first Ca^{2+} ligand in each domain (D452 in EGF12 and N490 in EGF13). N490 has a bulkier side chain (Box II) when compared to D452 (Box III) as indicated by arrows. The side

chain amide group of N490 points away from the Ca^{2+} because of its electropositive character. The side chain of N490 leads to an offset of both Y482 and I509. This offset has been described as a “gently curved” structure when the coordinates were published (Cordle et al., 2008a).

Measurements of interdomain tilt and twist angles (Figure 6.3) were performed using the software mod2 (Downing et al., 1996). Whereas the Ca^{2+} ligands were used as reference residues in the previous study, here the position of the two major β-strands are used as reference residues to allow measurements of angles in non Ca^{2+} binding domains. The domain twist/tilt angles are 119.3°/9.7° for the EGF11-12 interface, and 132.0° and 20.3° for the EGF12-13 interface, which is in line with the offset described in Figure 6.2.

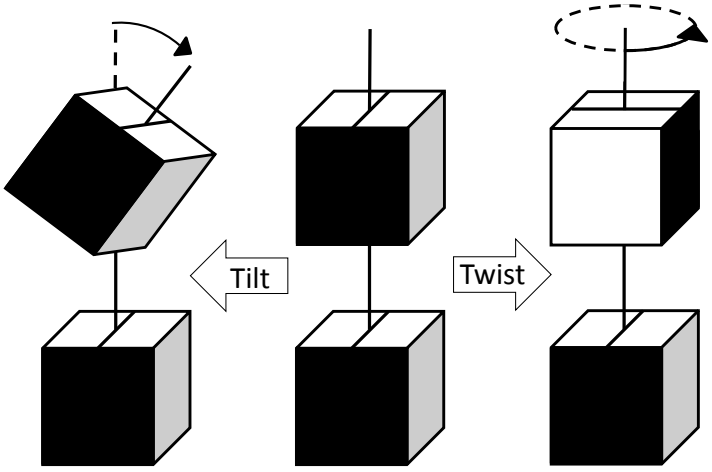


Figure 6.3 Interdomain tilt and twist angles. The relative position of two domains attached to each other can be described by a combination of tilt and twist angles.

When overlaying EGF11 (green, Figure 6.4 left) and EGF12 (red, Figure 6.4 left), the Ca^{2+} ions are offset to a much greater extent than observed for EGF12 and EGF13 (Figure 6.2) because of an additional carbonyl backbone ligand at S435 in EGF11. Although carbonyl backbone ligands in both Ca^{2+} binding sites of cbEGF22-TB4-cbEGF23 have been described (Lee et al., 2004), it is unclear whether this ligand arrangement is a property of EGF11 or due to the lack of the preceding domain EGF10 in this construct. However, if it is a property of EGF11, the

difference in Ca^{2+} position may result in significantly different tilt and twist angles for the EGF10-11 pair. This is investigated in more detail in chapter 8.

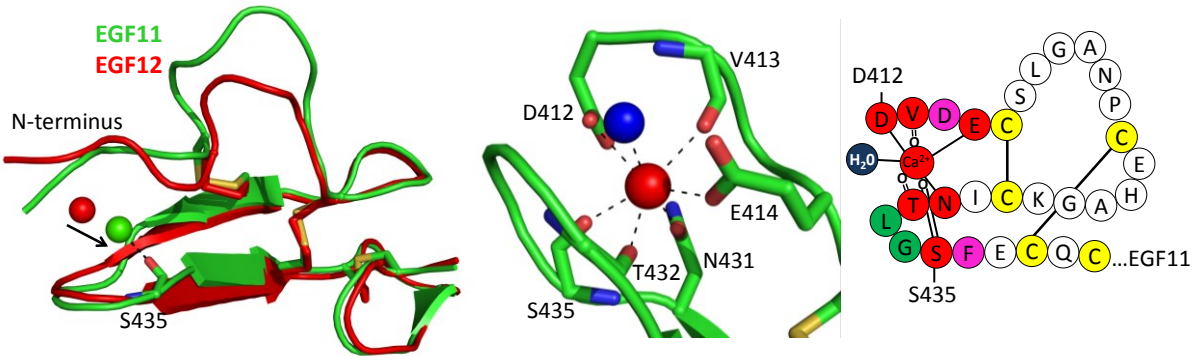


Figure 6.4 Ca^{2+} is offset due to an additional Ca^{2+} ligand in EGF11. S435 (EGF11) provides a carbonyl backbone ligand and hence pulls the Ca^{2+} ion more towards the second strand of the major β-sheet (arrow on the left, EGF11 in green, EGF12 in red). Ca^{2+} binding in EGF12 is described in section 4.1 (Figure 4.2). In EGF11 all other ligands are at identical positions and another water molecule (blue) provides the seventh ligand (middle and right picture). In the schematic diagram, calcium and calcium binding residues are depicted in red, H_2O in blue, interdomain packing residues in green, cysteines in yellow and conserved residues for intradomain stability in purple.

6.2.2 Phi/Psi torsion angle comparison for hN-$1_{11\text{-}13}$

The structure of a protein backbone can be described by a combination of the amino acid sequence (primary structure) together with the torsion backbone angles (Figure 6.5)

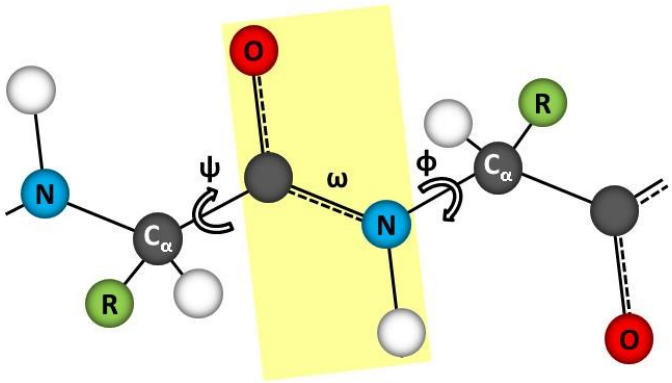


Figure 6.5 Torsion angles ϕ, ψ and ω of the protein backbone structure. The peptide bond has a partial double bond character, ω (C_α-C'-N-C_α) is usually 180° (*trans*), in a few cases the energetically less favourable angle 0° (*cis*) is adopted. The planar peptide bond is indicated in yellow. The angle ψ (N-C_α-C'-N) and the angle ϕ (C'-N-C_α-C') determine the backbone fold of the polypeptide. Note that bond length is not representative.

Structural characteristics of EGF domains are thus reflected in their torsion angles; an overlay of ψ, ϕ torsion angles of individual domains in hN-1_{11-13} is presented in Figure 6.6.

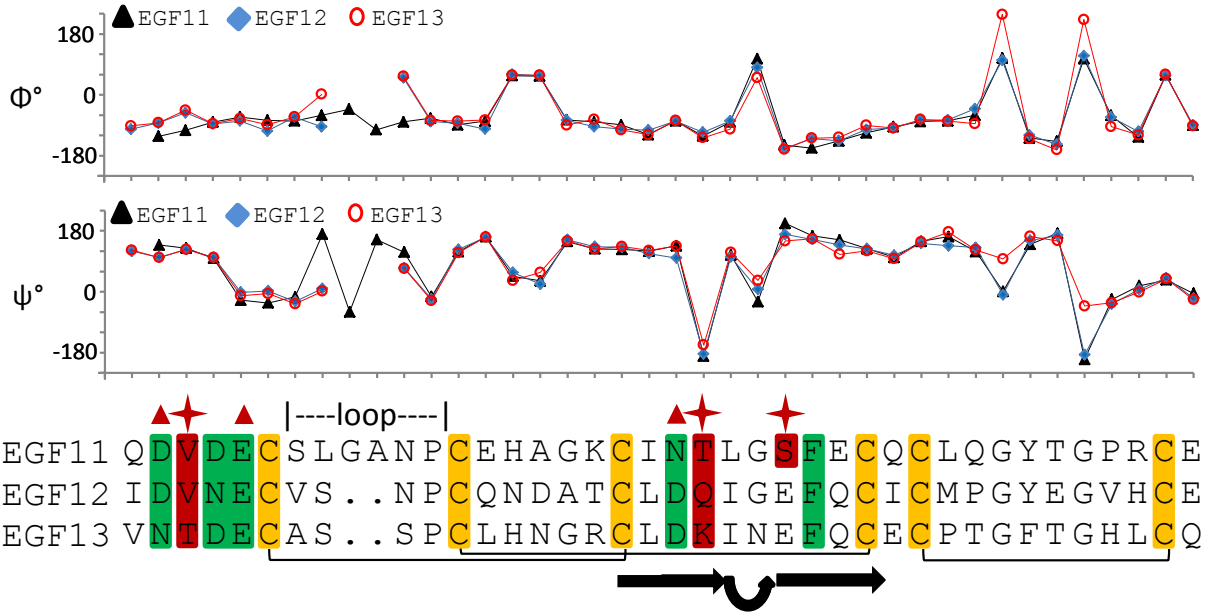


Figure 6.6 ψ/ϕ torsion angles derived from the hN-1_{11-13} crystal structure (Cordle et al., 2008a). The schematic diagram at the bottom shows the sequence alignment of individual domains. Consensus residues for Ca^{2+} binding are shown in green, cysteines in yellow. Ca^{2+} ligands are indicated in red, with a red triangle for side chain ligands and a red star for carbonyl backbone ligands. The position of the loop in between C1 and C2 is indicated as well as the position of the β-sheet (arrows). Most phi/psi angles are very similar among the EGF domains. The very first residues of EGF11 are different as EGF11 is the N-terminal domain. Two conserved glycines at the C-terminus of EGF13 also show different angles. Note that the N-terminal residue in EGF11 was not present in the crystal structure.

The overlay of ψ and ϕ angles clearly indicates the high conservation of backbone structure, a fact that agrees well with the overlays shown in Figure 6.2 and Figure 6.4. With respect to the differences described for EGF11-12 and EGF12-13, instead of a small number of distinctly different torsion angles at the C-terminus of EGF11 or EGF12 and the N-terminus of EGF12 and EGF13, small differences are observed for a wide range of residues at these positions in the sequence. This suggests that the differences in twist/tilt angles of these domain pairs cannot be easily described by significant differences in torsion angles at the interface. It is more likely, that the difference in tilt and twist angles is balanced over many residues. N-terminal residues of EGF11 show slightly different torsion angles when compared to EGF12 and EGF13. It is not clear

whether this is due to the difference in Ca^{2+} positioning described (Figure 6.4) or due to the non-native character of EGF11 in this construct. Similarly, EGF13 as a C-terminal domain shows different torsion angles for two conserved glycines between C5 and C6 when compared to EGF11 and EGF12. The differences in phi/psi angles of these glycines may suggest a role during the formation of the interdomain packing interaction or could be the result of the non-native birA-tag at the C-terminus of this construct, which is important in rigidifying this region (Cordle et al., 2008a). Angles for residues belonging to the interdomain region are consistent for EGF11/12 and for EGF12/13. Angles are different in the loop of variable length in between C1 and C2. Small differences are observed in the β-sheet region and in EGF11 around the S435 Ca^{2+} ligand.

6.3 Crystallography

Crystal trays were set-up in collaboration with Professor Susan Lea at the Sir William Dunn School of Pathology. Trays for constructs hN-1_{4-7}, hN-1_{7-9}, hN-1_{8-11}, hN-1_{9-11}, hN-1_{11-13}12DG, hN-1_{21-23} and hN-1_{20-23} were set up at different conditions and concentrations of Ca^{2+} and Ba^{2+} (1-10 mM). Crystals were successfully grown for constructs hN-1_{4-7}, hN-1_{21-23} and for hN-$1_{20-23;}$ Figure 6.7 shows images of hN-1_{20-23} crystals.

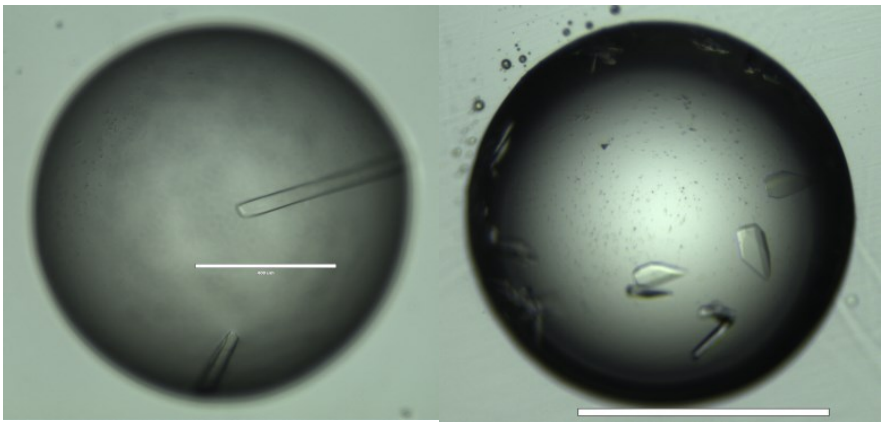


Figure 6.7 hN-1_{20-23} crystals. On the left, crystals were grown with a buffer containing 0.1 M Na HEPES pH 7.0, 10% w/v PEG 6000. On the right, buffer conditions were 0.1 M sodium acetate, pH 5.5, 12% PEG w/v 5000 MME. The bar represents 400 µm.

The data set used for the hN-1_{21-23} was crystallised in 0.2 M ammonium chloride 20% PEG 3350. For hN-1_{4-7}, crystals were grown in 0.15 M potassium bromide 30% w/v PEG 2000 MME and in 0.2 M potassium formate, 20% w/v PEG 3350 (JSCG screen).

Figure 6.8 shows an overview of crystal structures obtained during the course of this thesis, which were solved by Professor Susan Lea's group, using protein produced and purified as described in chapter 2 and 3. Crystal structures were obtained for regions flanking EGF4 and EGF22, but not for EGF10. The region around EGF22 was crystallised in the hN-1_{21-23} and hN-1_{20-23} constructs, and EGF6 in the hN-1_{4-7} construct. The crystal structure of hN-1_{4-7} was solved by molecular replacement with the factor IX coordinates (Rao et al., 1995) at 2.4 Å resolution. The hN-1_{21-23} and hN-1_{20-23} crystal structures were solved using the heavy atom Ba^{2+} (multiple isomorphous replacement) at 1.5 Å and 1.7 Å resolution, respectively.

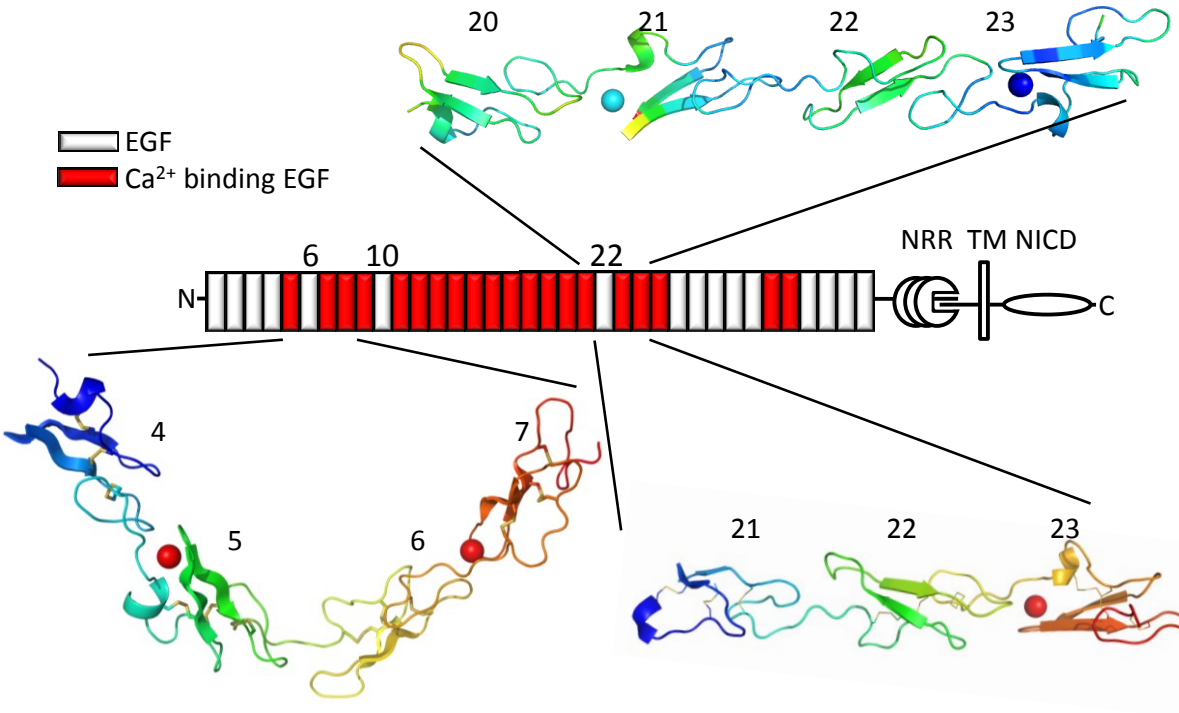


Figure 6.8 Crystal structures of hN-1 obtained during the course of this thesis. Note that there was no Ba^{2+} in EGF21 of the hN-1_{21-23} crystal structure, most likely due to its low affinity.

6.4 Domain orientation

Interdomain tilt and twist angles were calculated using the program mod2 (Table 6.1) as described in section 6.2. In addition to calculating the twist and tilt and angles between the two highly conserved β-strands, the position of the six conserved cysteine residues was used as well. Small differences were observed between the two strategies applied, demonstrating that the accurate measurement of tilt and twist angles is not straightforward. However, these differences are smaller than the differences seen for the twist/tilt angles for cbEGF-cbEGF, EGF-cbEGF and cbEGF-EGF pairs.

Table 6.1 Interdomain tilt and twist angles for EGF domains in hN-1. Three different classes of interfaces were analysed, cbEGF-cbEGF interfaces, EGF-cbEGF interfaces and cbEGF-EGF interfaces.

		β-sheet overlay		cysteine overlay	
context	**domains**	**twist (°)**	**tilt (°)**	**twist (°)**	**tilt (°)**
cbEGF-cbEGF	**11-12**	119.3	9.7	126.7	8.1
	12-13	132.0	20.3	132.3	15.9
EGF-cbEGF	**4-5**	179.7	32.8	173.2	34.0
	6-7	151.0	38.6	159.3	42.2
	22-23	171.4	42.6	165.1	48.2
cbEGF-EGF	**5-6**	112.6	90.3	119.5	87.6
	21-22	51.4	4.5	56.5	0.4

The EGF-cbEGF interfaces 4-5, 6-7 and 22-23 have a significantly larger twist angle than the cbEGF-cbEGF interfaces described for the domain pairs 11-12 and 12-13, and also the tilt angle is larger. Larger twist angles had been described before in the fibrillin-1 cbEGF-cbEGF domain pairs 11-12 and 32-33; these are type I EGF domains and have one residue less in the linker region (Smallridge et al., 2003, Downing et al., 1996). Angles observed for the two cbEGF-EGF domain pairs 5-6 and 21-22 were different, the 5-6 interface has a very different tilt angle of 89.9°, but the twist angle of 112.4° is not too different from that observed for cbEGF-cbEGF domains. In contrast, the 21-22 interface has a tilt angle of 4.5° which was just below those observed for the 11-12 and 12-13, but the twist angle of 51.4° was significantly different. The construct hN-$1_{8\text{-}11}$

did not crystallise, nor did hN-1_{9-11}. This could be due to the fact that EGF9 was the only EGF domain without the conserved aromatic packing residue (Table 1.1), which could introduce flexibility within the EGF9-10 interface. These results suggest minor differences in the interfaces formed by cbEGF-cbEGF and EGF-cbEGF domain pairs, and major differences among the interfaces formed by cbEGF-EGF domain pairs.

6.5 Structural analysis of hN-1_{4-7}

6.5.1 Overview of structures

Crystal structures of two crystal isoforms were obtained for hN-1_{4-7}. One of them had two molecules in the asymmetric unit (Figure 6.9).

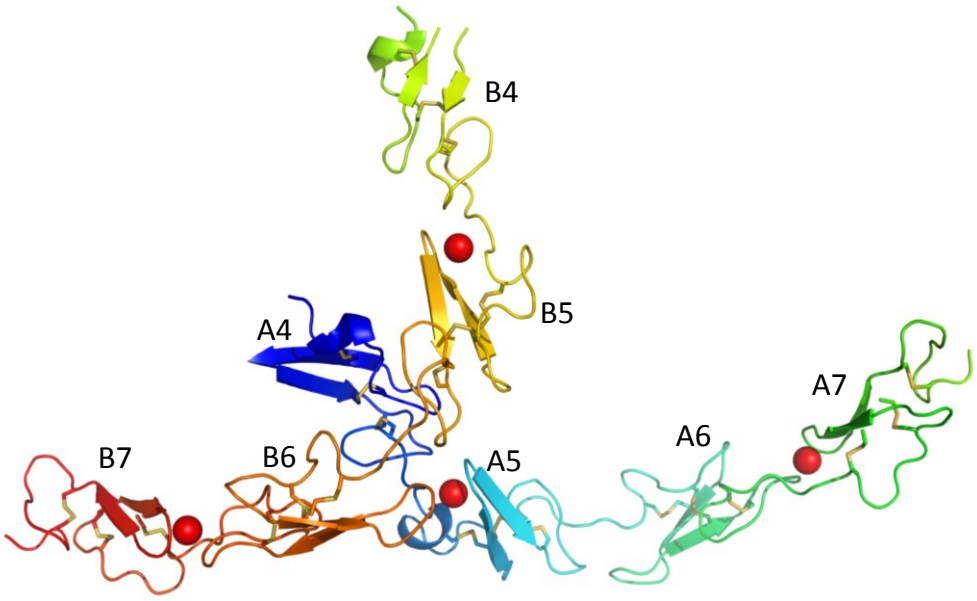


Figure 6.9 Unit cell with two molecules A and B of construct hN-1_{4-7} (space group P1211). The molecules interact at the interface of EGF5-6 in molecule B and at the EGF4-5 interface of molecule A.

The packing of the molecules in the unit cell and the unusual tilt angle of around 90° in between EGF5-EGF6 lead to the question of whether this angle is due to crystal packing. A second crystal that was solved has a single unit in the asymmetrical unit and space group C121 and showed a similar interdomain tilt angle. The align function of Pymol was used to overlay the

EGF-5 domain of all three solved structures. Whereas the peptide backbone and Ca^{2+} of EGF5 overlayed nearly perfectly in all three crystal structures, all three domain interfaces show small variations (Figure 6.10). The EGF5-6 interface, for example, is similar for P1211A and C121. However, the P1211 B isoform shows a slightly wider angle; this could potentially be explained by crystal packing within the unit cell since the EGF5-6 interface of P1211B packs with the P1211A EGF4-5 interface.

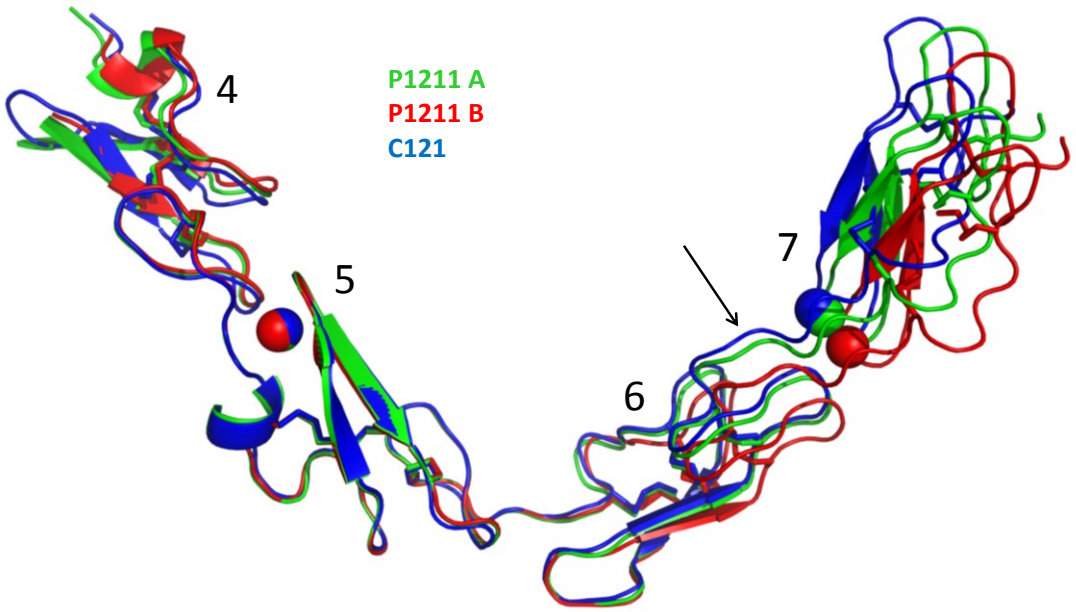


Figure 6.10 Overlay of three solved crystal structures of hN-1_{4-7}. The Pymol function align was used to overlay residues belonging to EGF5 of molecule A in the crystal structure with the space group P1211 (green), molecule B (red) and of another crystal structure with the space group C121 (blue). Note that although interdomain angles are similar, EGF domains are not overlaying perfectly. A small difference can for example be observed on the EGF5-6 interface which is similar for P1211A and C121, but slightly different for P1211B.

Interdomain twist and tilt angles were calculated using the program mod2 (Table 6.2). Differences in twist and tilt angles observed between the individual crystal structures correlate well with the observations in Figure 6.10. For example, a slightly larger tilt angle is shown for the EGF5-EGF6 interface for the C121 crystal form as indicated by an arrow. Differences of 5-10° in twist angles are also observed between the three structures. These differences may arise from differences in crystal packing of molecules.

Table 6.2 Interdomain tilt and twist angles for EGF domains in hN-1$_{4-7}$. Three different crystal structures were analysed that showed small variations among the interfaces.

	β-sheet overlay (°)		cysteine overlay (°)	
crystal	**twist**	**tilt**	**twist**	**tilt**
4-5 P1211 A	179.7	32.8	173.2	34.0
4-5 P1211 B	179.7	33.2	187.9	31.0
4-5 C121	186.6	31.2	182.2	30.8
5-6 P1211 A	112.6	90.3	119.5	87.6
5-6 P1211 B	122.92	83.1	128.6	79.5
5-6 C121	113.43	90.0	120.3	89.8
6-7 P1211 A	151.0	38.6	159.3	42.2
6-7 P1211 B	150.7	46.3	160.3	49.5
6-7 C121	144.8	49.9	153.6	50.3

Figure 6.11 shows an example for packing of the crystal solved in the space group C121. The region between C6 in EGF5 and C1 in EGF6 is shown in spheres and closely packs with that of another molecule (A).

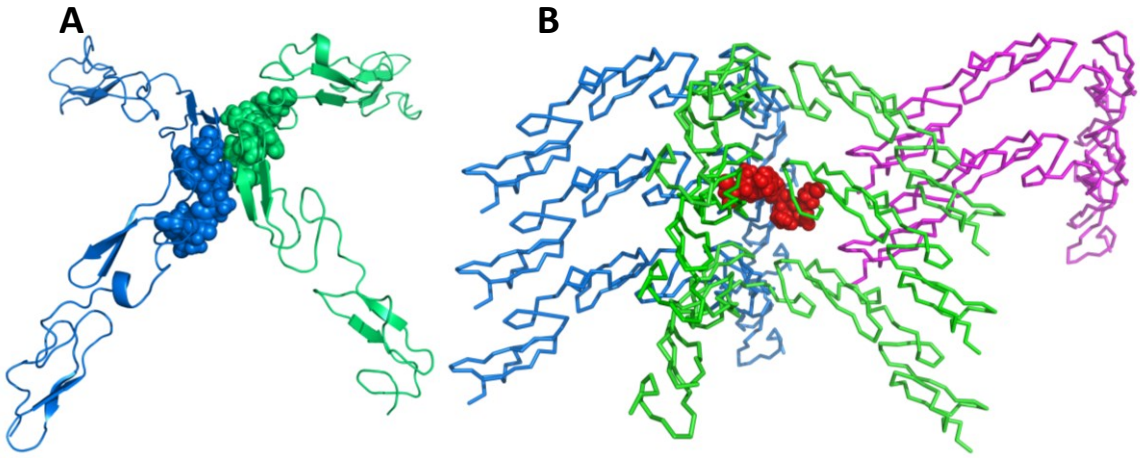


Figure 6.11 Crystal packing of hN-1$_{4-7}$ in the C121 space group. Spheres in A and B indicate the region between C6 in EGF5 and C1 in EGF6. In (A), two molecules are shown that pack around the interface of EGF5-6. In B, all molecules within 20 Å of the EGF5-6 interface (red) are shown. The crystal packing may explain certain differences in interdomain angles observed.

In (B) all molecules within a distance of 20 Å of the EGF5-6 interface are shown, parallel molecules share the same colour. It is not clear how much the twist and tilt angles of the interfaces formed are affected by crystal contacts.

Similar to other EGF domain structures, little secondary structure was found in the hN-1_{4-7} construct. However, all four EGF domains showed the conserved major and minor β-sheets that were described earlier (Bork and Campbell), as well as conserved cysteines and disulphide bonds. Figure 6.12 shows an unusual short α-helix at the beginning of EGF5.

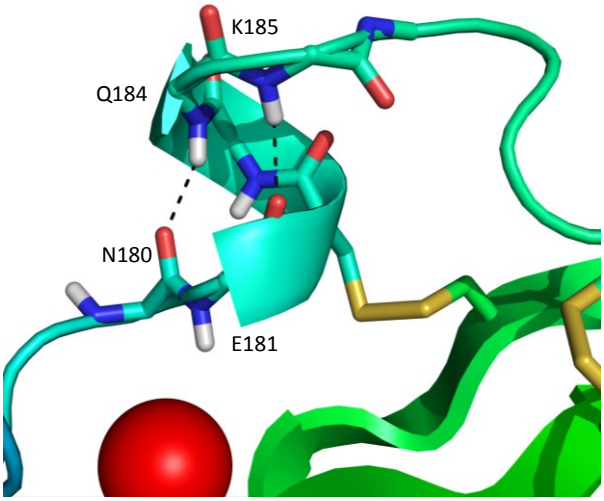


Figure 6.12 Short α-helical sequence at the beginning of EGF5. A similar, but less defined helix was seen at the equivalent position in EGF4.

Figure 6.13 shows an analysis of phi/psi torsion angles observed for individual domains of the three crystal structures solved, and boxes highlight regions were differences are observed. Apart from the residues in between C1 and C2 which are usually less well defined, such as this region in EGF5 (C182-C189), most phi/psi angles agree well among the different crystal structures. For the residues at the turn in major β-sheet in EGF4, no crystal coordinates were obtained due to the lack of electron density. This region is the consensus packing region to the preceding domain EGF3 and is probably not structured because of the lack of EGF3. With respect to the interface formed in between EGF5-EGF6, only small differences are observed at the end of EGF5 and the beginning of EGF6. However, a close up of the three structures (Figure 6.14) indicates a deviation at three points, the first around P221 and the other two at both strands of the major β-sheet in the region G237-T240. Residues P221, T236 and T240 are highlighted in Figure 6.13.

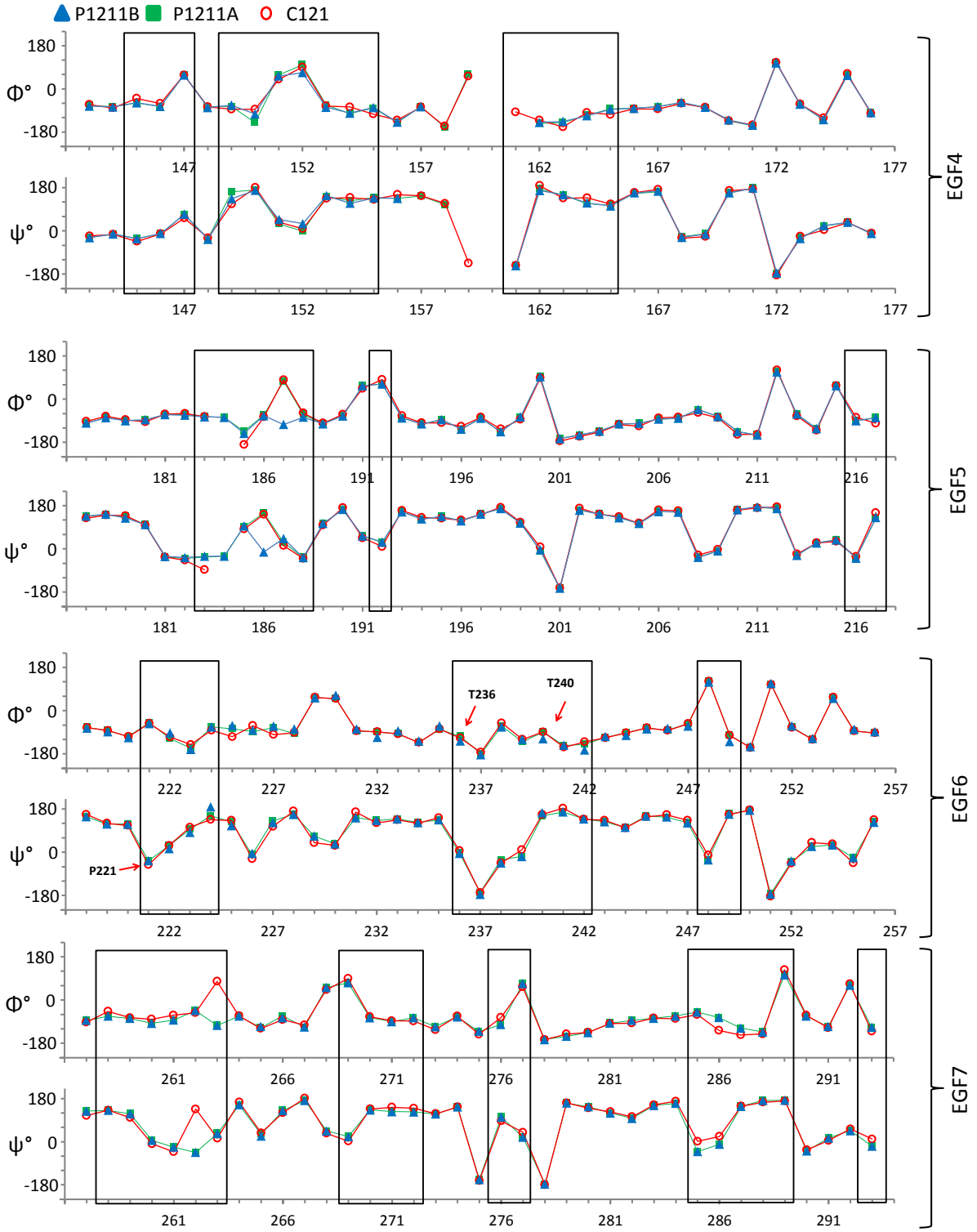


Figure 6.13 Overlay of phi/psi torsion angles in hN-1$_{4-7}$. Boxes highlight individual regions where phi/psi torsion angles vary slightly. These are mostly the residues between C1 and C2 (left hand boxes), but smaller differences are observed across the whole molecule. Note that the residues 157-161 were not well defined in the electron density, probably because of the lack of packing against N-terminal domain EGF-3. Note the subtle differences in residues 221/222 and 238/239 and 241/242 that show a deviation in the EGF5-6 interface (Figure 6.14).

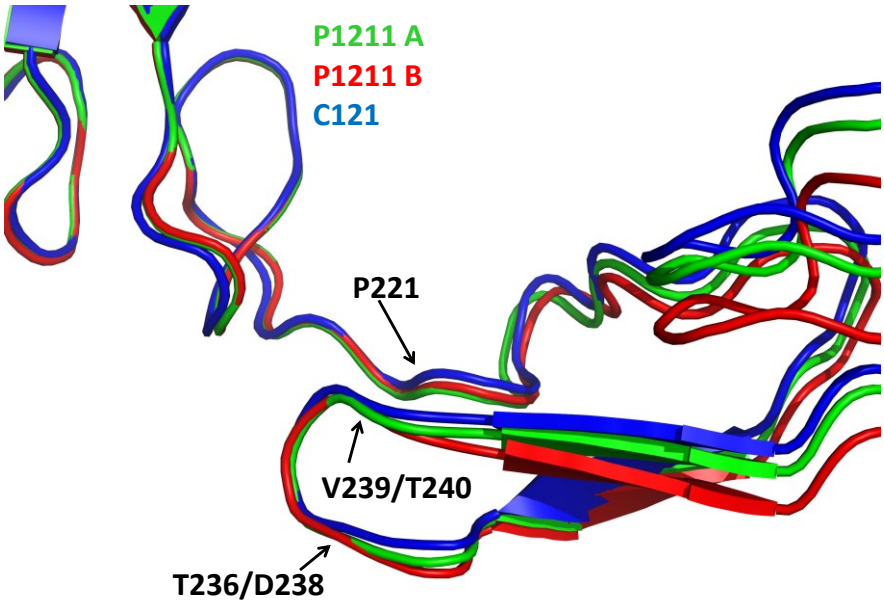


Figure 6.14 Deviation in the EGF5-EGF6 interface. The backbone strands at the EGF5-EGF6 interface appear to deviate at three positions in the backbone structure. This is around P221 and on both strands of the major β-sheet (T236-T240).

Figure 6.15 shows phi/psi torsion angles of individual EGF domains in the hN-1_{4-7} crystal structures compared to torsion angles of EGF12 in the hN-1_{11-13} crystal structure. Similar to hN-1_{11-13}, torsion angles are mostly consistent with the exception of the loop region. For example, the presence of a *cis*-proline (P224) causes a huge deviation of torsion angles in this region. However, a larger variation is also seen in the region around the turn of the major β-sheet. EGF6 has an additional residue in between C3 and C4 and shows different torsion angles. Whereas EGF13 showed differences in the glycines 3 and 6 residues after C5 when compared to EGF11 and EGF12, here, differences are seen in the second and third residues after C5 that showed similarity for EGF6 and EGF12 and for EGF4, EGF5 and EGF7. This correlates with the presence of a glycine three residues after C5 in EGF6/12, but not in the other domains. Residues after this position showed very similar angles. The very first residues showed little difference too, and the fact that EGF6 is not a Ca^{2+} binding domain was not evident from this plot.

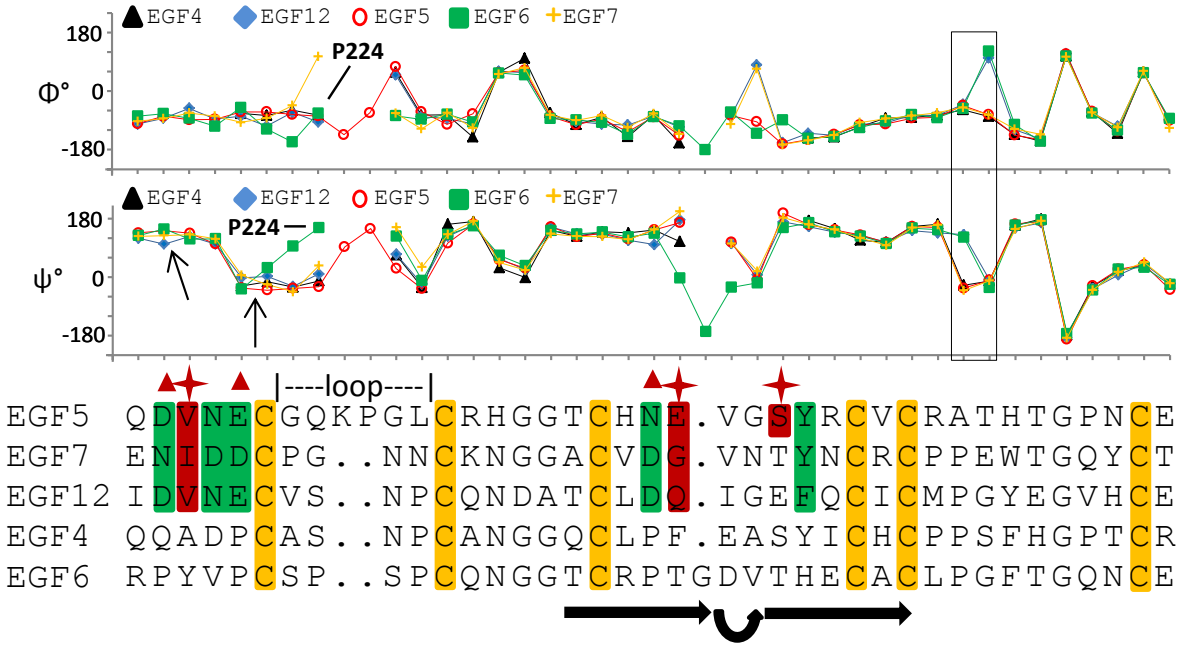


Figure 6.15 Torsion angles of individual EGF domains of hN-1$_{4-7}$ and EGF12 of the hN-1$_{11-13}$ crystal structure. Torsion angles are highly consistent for several regions. However, the presence of a glycine three residues after the fifth cysteine in EGF6 and EGF12 seems to have a large effect on its own phi angle and on the psi angle of the previous residue (box). Arrows on the left indicate critical regions for the EGF5-6 interface (section 6.5.2), the *cis*-proline P224 (EGF6) is indicated. Conserved cysteines are highlighted in yellow and used as a reference for the alignment. Consensus residues for Ca^{2+} binding in green, the red triangle indicates the donation of a Ca^{2+} side chain ligand. Highlighted in red and with a red asterisk are residues that donate a carbonyl backbone Ca^{2+} ligand. The arrows underneath the sequence alignment indicate the position of the beta-sheet.

6.5.2 The EGF5-EGF6 interface

Figure 6.16 shows residues involved in the EGF5-6 interface. In the position of the conserved aromatic packing residue, there is a histidine in EGF5 (H210) that packs with the residue at position 2 after C2 in EGF5, which here is H191; this interaction is similar to that observed in other EGF domains (section 1.4.4). However, H210 does not pack with the turn of the major β-sheet in EGF6 or any other residue within EGF6, in contrast to the interaction typically observed in cbEGF pairs. The packing residue here is A208, and to a smaller extent T209 (not shown). The corresponding packing region, the turn in the major β-sheet in EGF6 is different as well to previously analysed EGF domains since it has an uneven number of residues between C3 and C4 (9 residues). Usually, there are 8 residues and the hydrophobic packing residue is at

position 4 after the third conserved cysteine in the EGF domain. Here, the residue V239 of EGF6 that packs with A208 of EGF5 is at position 6 after the third conserved cysteine in EGF6. G237 and D238 have no contact to EGF5.

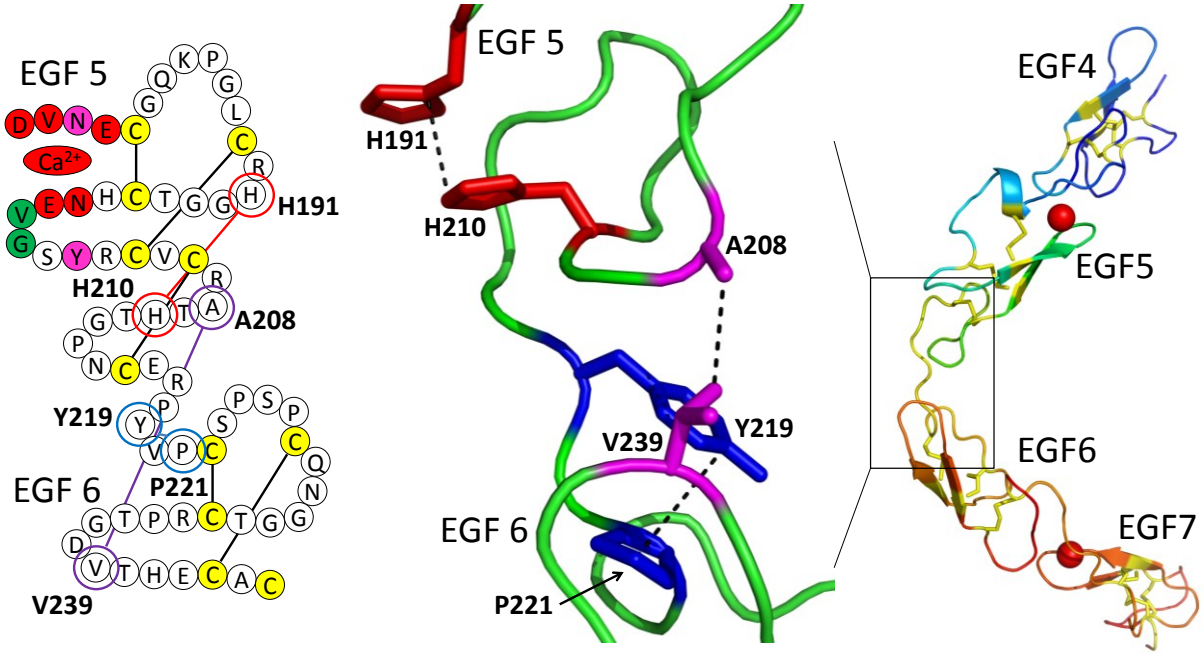


Figure 6.16 Interface between EGF5 and EGF6. The bend observed between EGF5-6 with a tilt angle of around 90° (crystal structure on the right) is due to the fact that packing involves A208 which is two residues before the conserved aromatic residue H210 and packs with V239 (purple). H210 packs, similar to what was observed earlier with H191 (red). Y219 packs against P221 (blue). In the schematic diagram on the left, Ca^{2+} and Ca^{2+} binding ligands are depicted in red, packing residues in green, cysteines in yellow and conserved residues for intradomain stability in purple.

Several structural features are identical to the ones observed for cbEGF domain contacts which were described in section 1.4.4 (not shown here). These are the packing of the side chain of T211 against the side chains of R217 and Y219. And also similar is the minor β-sheet between T211 and R217. However, Y219 seems to play a crucial role as well in this interface since it further packs with P221 which is stabilised by hydrophobic contacts to surrounding regions of the domain such as P235 within the major β-sheet.

The fact that the conserved residue H210 does not pack with the turn in the major β-sheet raises several questions. Is a histidine able to replace the conserved aromatic residue at this

position? No other domain in human Notch-1 has a histidine at this position (table 1.1, section 1.4.4). In EGF14 of fibrillin-1 calcium chelation experiments indicated a high affinity (~7 µM) indicative of interdomain packing to EGF13 that has a histidine as the packing residue with EGF14 (Lee et al., 2004). However, the packing region between C5 and C6 is significantly longer in fibrillin-1 (12 residues in EGF13) compared to 8 residues in EGF6 of Notch-1 and no structural data are available to confirm that this histidine is the packing residue.

An overlay of EGF5 on EGF11 of the domain pairs EGF5-EGF6 and EGF11-EGF12 is shown in Figure 6.17. Both domains are very similar and deviate just after the linker region, around P218 (EGF6) which is a very atypical position for a proline. This has a very different torsion angle than the corresponding residue in EGF12.

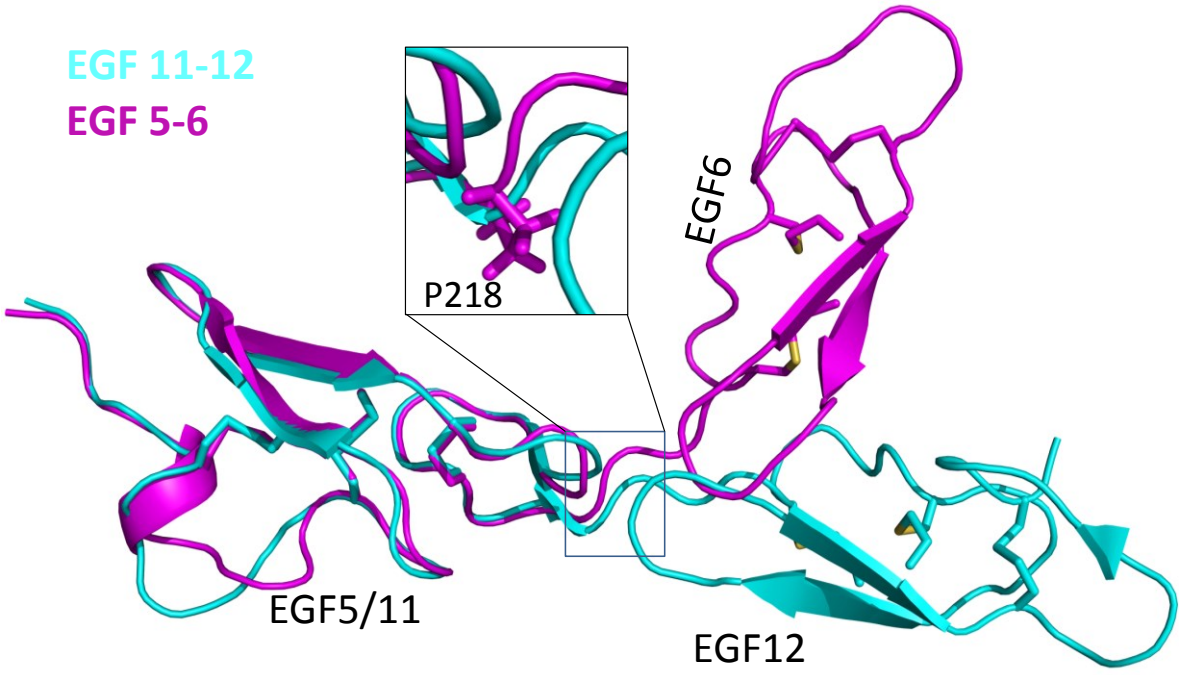


Figure 6.17 Overlay of EGF 5 on EGF11 of the domain pairs EGF5-6 (magenta) and EGF11-12 (cyan) in hN-1. Both domains are highly conserved in structure, but deviate at the position of P218 at the beginning of EGF6. P218 has phi/psi torsion angles of -70°/148°, whereas, the corresponding residue in EGF12 (D452) has phi/psi torsion angles of -84°/ 102°.

Figure 6.18 shows the complementary alignment with EGF6 on EGF12. Here too, the deviation starts at a proline (P221) which is also at an atypical position. It needs to be noted that a *cis*-proline in the loop between C1 and C2, P224, leads to a further local change in the backbone structure. However, the backbone and side chains of the C1 atom are not very different in the overlay of both constructs most likely because of the conserved disulphide bond. The *cis*-proline thus affects mainly residues within the loop only, but also leads to an offset of the direction of the peptide chain around C222.

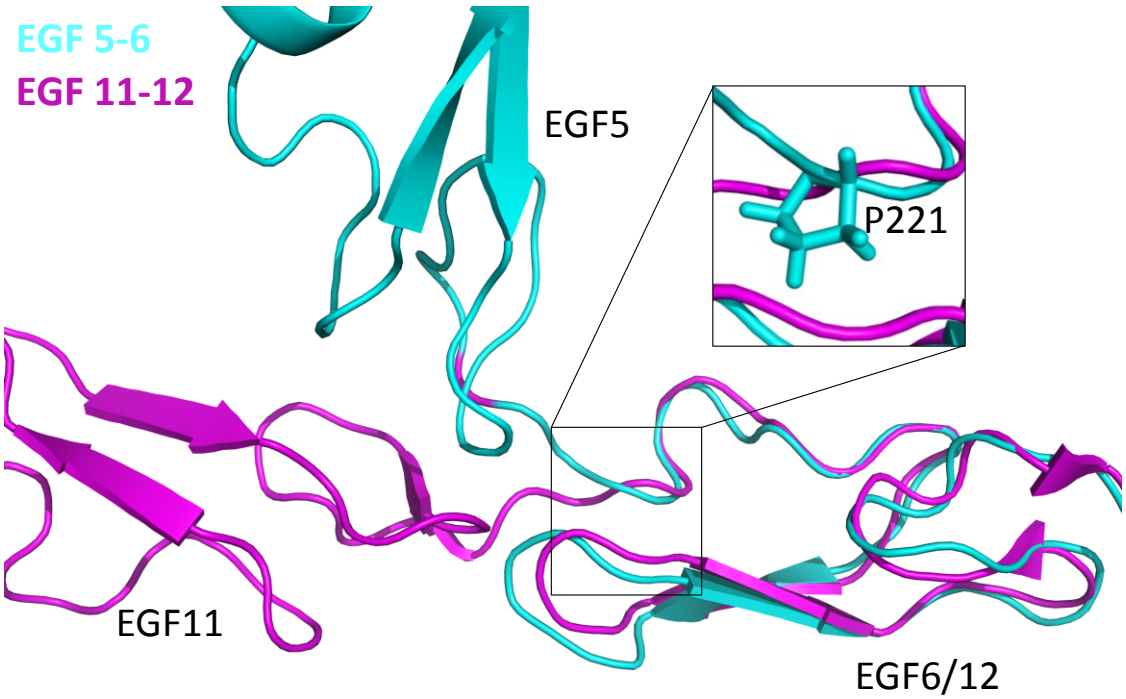


Figure 6.18 Overlay of EGF 6 on EGF12 of the domain pairs EGF5-6 (cyan) and EGF11-12 (magenta). The chain starts to deviate from P221 in EGF6. C1 (C222) has a slightly variation of the backbone position because of an atypical *cis*-proline (P224). The phi/psi torsion angles are -50°/-36° for P221 which is different to -77°/-2° for E455 in EGF12.

6.5.2.1 NOESY analysis of the EGF5-6 interface

In chapter 4, NOESY experiments were performed to describe the packing of the aromatic consensus residue with the β-turn of the following domain. Here, 2D NOESY experiments on hN-$1_{4\text{-}7}$ and 3D NOESY experiments on hN-$1_{5\text{-}7}$ and hN-$1_{4\text{-}6}$ indicate that the packing residue A208 is in spatial proximity to V239, demonstrating that the bend observed in the crystal structure is

probably not due to crystal packing effects and exists in solution. Figure 6.19 shows the corresponding strip plots. This interface and the interdomain twist and tilt angles present in solution are analysed further in chapter 8.

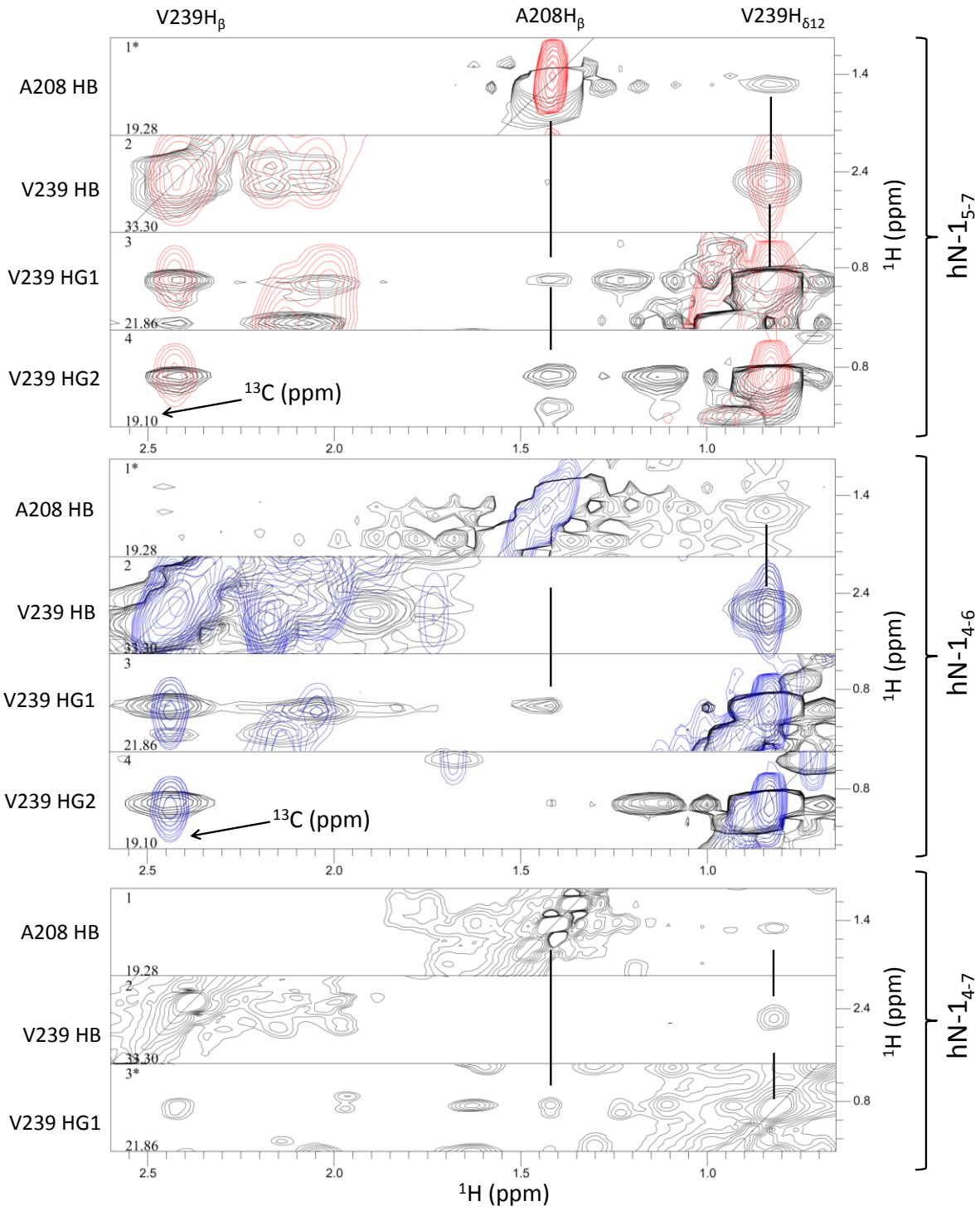


Figure 6.19 Strip plots of A208 and V239 indicate NOESY contacts. These were observed in all three constructs tested, hN-$1_{5\text{-}7}$ (top), hN-$1_{4\text{-}6}$ (middle) and hN-$1_{4\text{-}7}$ (bottom). 3D HCCH-TOCSY peaks are shown in red (hN-$1_{5\text{-}7}$) and blue (hN-$1_{4\text{-}6}$), 3D ^{13}C NOESY (hN-$1_{5\text{-}7}$ and hN-$1_{4\text{-}6}$) and 2D ^{1}H NOESY (hN-$1_{4\text{-}7}$) in black.

6.5.2.2 EGF5-EGF6 cell aggregation assay

The physiological role of the EGF5-6 bend is unclear and raises the question of whether this bend is conserved or not. In *Drosophila*, Y219 and P221 are conserved (Figure 6.20), but P218 is not and is a lysine. However, lysine has similar predicted phi/psi torsion angle preferences (not shown). Since Y219 and P221 are conserved in *Drosophila*, the packing formed by these two residues was chosen for mutagenesis. Three constructs were designed. The insertion of 3 glycines is expected to disrupt packing and introduce flexibility; effects on packing are expected from a Y->A substitution. The insertion of 3 prolines is expected to disrupt the packing between Y219 and P221, without the introduction of flexibility.

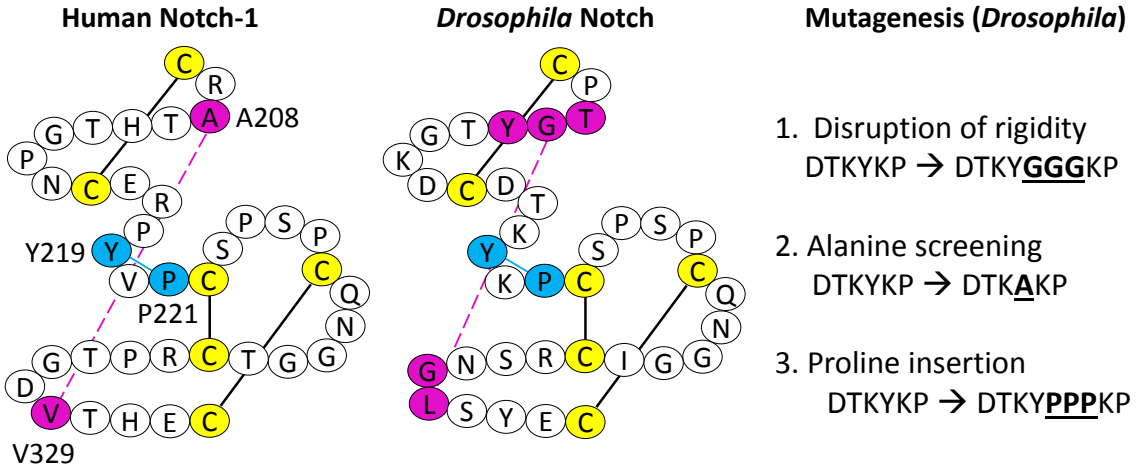


Figure 6.20 Alignment of hN-1 (left) and *Drosophila* Notch (middle) of the EGF5-6 interface. On the right, the mutagenesis of the *Drosophila* protein sequence is described. Note that it is not clear which residues provide the interdomain packing in *Drosophila*, but candidates for packing are indicated in purple. Conserved cysteines are depicted in yellow, residues Y219 and P221 in blue.

The introduced mutations could have an effect on ligand binding or on subsequent mechanisms such as the endocytosis by the ligand-expressing cell. Notch mutations were introduced in full-length Notch heat-shock vector and transfected into S2 cells using the calcium phosphate method (section 2.13). A Serrate construct in pUAST, with GFP attached to the intracellular domain (kind gift of Paul Taylor), was simultaneously transfected in a separate well.

The transfection efficiency was tested by staining all cells with DAPI, and an antibody against the intracellular domain of Notch. The intrinsic fluorescence of GFP was sufficient to assess Serrate expression. The number of cells transfected was assessed under the fluorescence microscope. Very low transfection rates were observed (~5%) and varied between the different transfections (not shown). Serrate cells were combined with cells transfected with different Notch constructs. Figure 6.21 shows a typical staining pattern obtained, here for the Notch wildtype (red) to Serrate (green) interaction. All nuclei are stained using DAPI (blue).

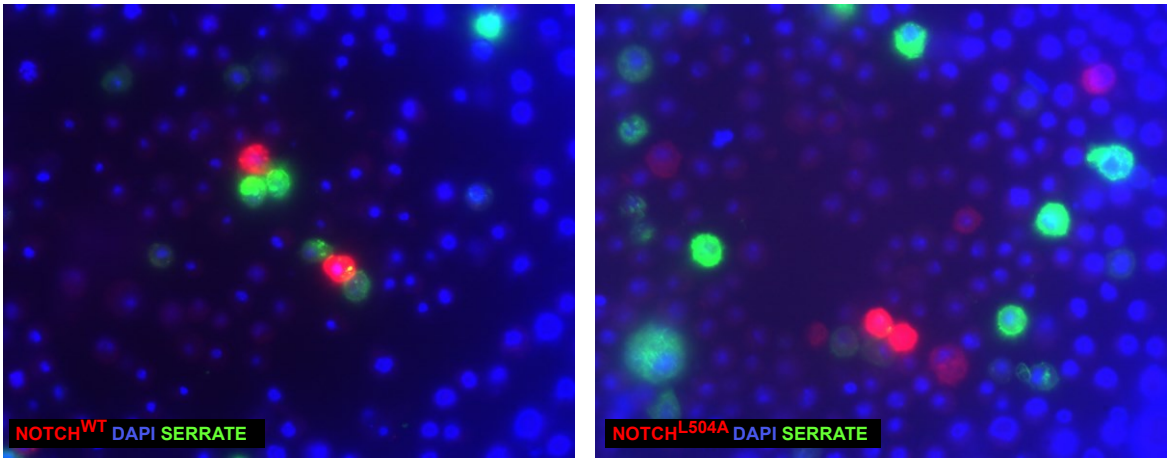


Figure 6.21 Immunofluorescence staining of wildtype Notch (left) and the Notch L504A mutant (right). Whereas Notch wildtype expressing cells were attached to Serrate expressing cells (left), many Notch L504A mutant transfected cells did not attach (right).

Serrate cells were counted as per cent attached to Notch expressing cells. Errors were estimated from three independent counts of at least 100 cells. The Notch mutant L504A (EGF12) was used as a negative control since it had been identified in previous cell aggregation assays as a key ligand binding residue (Beatriz Hernandez de Madrid, PhD thesis, University of Manchester). Whereas immunofluorescence staining did not show evidently that L504A is less well presented on the cell-surface, a clear reduction in cell attachment when compared to the wildtype was observed (Figure 6.22). Analysis of the different Notch mutants did not show a reduction in attachment. The experiment was repeated and showed little difference in the numbers obtained.

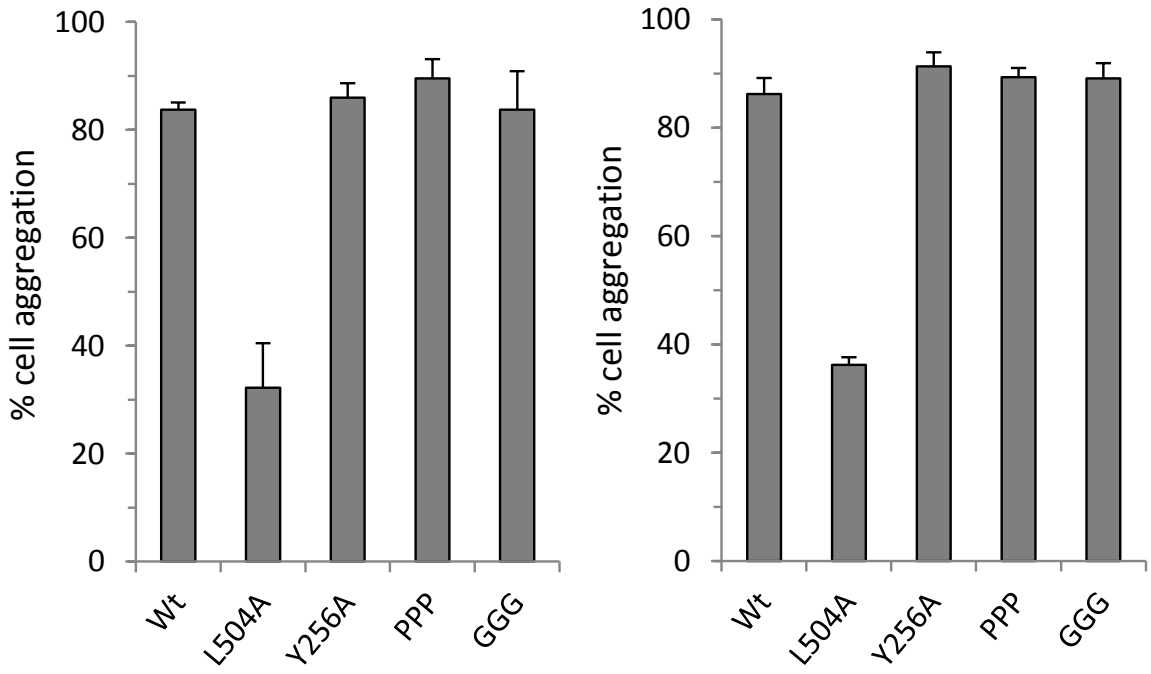


Figure 6.22 Cell aggregation assay with different Notch mutants. The graphs show the percentage of total Serrate expressing cells bound to cells that express different Notch constructs. The whole experiment was repeated with independent transfections (left and right). A consistent reduction in binding to the L504A mutant, but not for the mutations introduced in the EGF5-EGF6 interface, was observed.

Consistent with the observation that EGF11-12 are required and sufficient to mediate ligand binding in the cell aggregation assay (Rebay et al., 1991), no effect was observed with mutations in the EGF5-EGF6 interface. Thus, the role of the bend in between EGF5 and EGF6 remains unclear. Since it has been shown that EGF11-12 are not sufficient for Notch signalling, further insights could be gained by testing various Notch fragments for binding against Serrate/Delta in FACS assays/SPR. Furthermore, a longer construct of Notch could be expressed in a eukaryotic system and the affinities of the mutants on the EGF5-EGF6 interface measured. Signalling assays could reveal whether the bend in between EGF5 and EGF6 and surrounding regions are necessary for the signalling activity of the receptor.

6.5.3 Ca^{2+} binding of EGF5 and EGF7

Figure 6.23 shows the structure of the Ca^{2+} binding sites in EGF5 and EGF7. In EGF5, Ca^{2+} ligands are carbonyls of V179, S201, and E198 and side chain ligands from D178, E181, and N197 and one ligand from a water molecule held in place by S169 (EGF4).

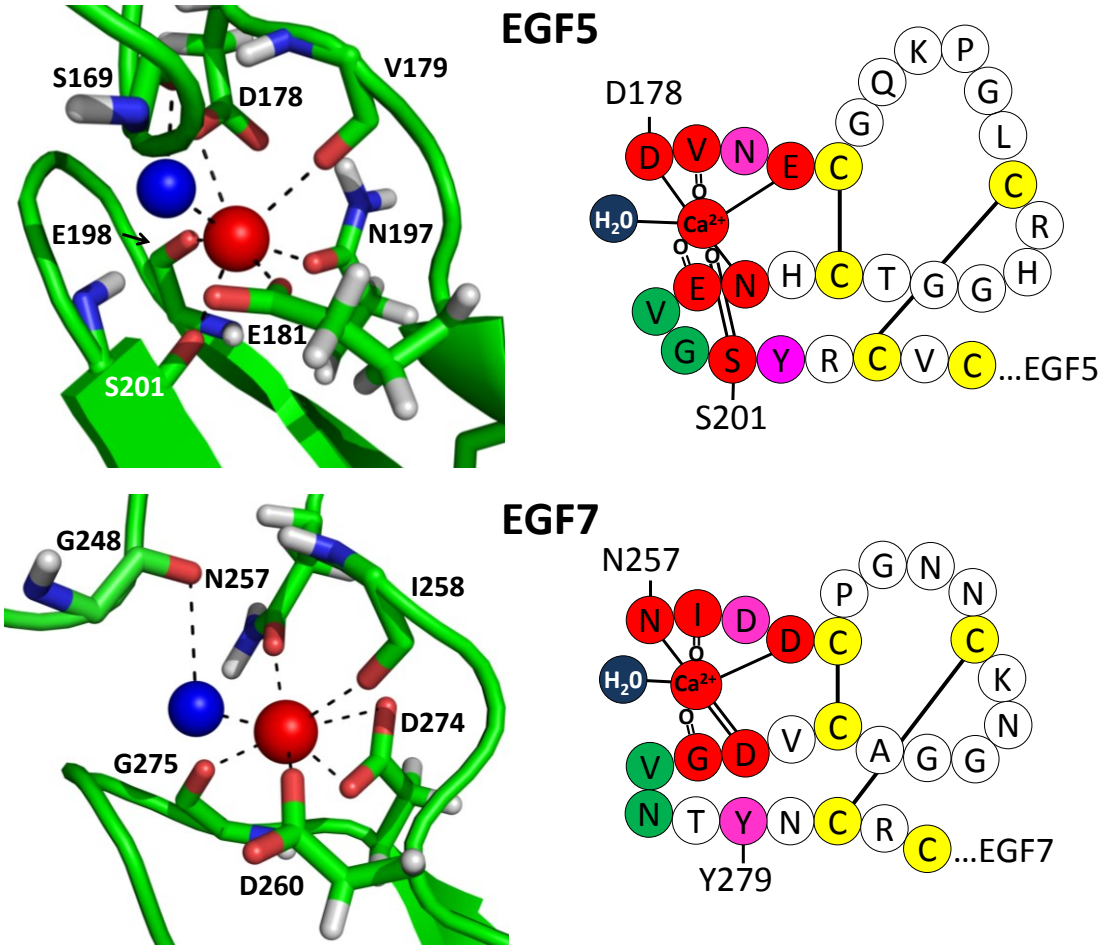


Figure 6.23 Ca^{2+} binding in EGF5 (top) and EGF7 (bottom) (C121). All seven ligands (red) including one water molecule (blue sphere) for both EGF domains were observed. In EGF5, S201 provides a backbone carbonyl ligand, which was seen previously in EGF11 of the hN-$1_{11\text{-}13}$ crystal structure. In EGF 7, two side chain ligands were provided by D274. In the schematic diagrams, calcium and ligands are depicted in red, cysteines in yellow, interdomain packing residues in green, water in blue and conserved residues for intradomain stability in purple.

A Ca^{2+} ligand at a position homologous to position S201 has been reported in EGF11 of the hN-$1_{11\text{-}13}$ structure (Cordle et al., 2008a), but also in the cbEGF22-TB4-cbEGF23 construct of fibrillin-1 (Lee et al., 2004). In EGF7, I258 and G275 provide backbone carbonyl ligands, whereas

N257 and D260 provide one side chain oxygen and D274 provides 2 side chain oxygens. Aspartic acid at the position of D260 is uncommon; at this position, a Ca^{2+} ligand has been shown to be donated by the side chain of either a glutamic acid or glutamine in an analysis performed in fibrillin-1 (Handford et al., 1991). T278 of EGF7 is not a Ca^{2+} ligand. EGF5 is in a native context in the hN-1_{4-7} construct and it shows the same ligand arrangement that was seen in EGF11 with a carbonyl ligand from a serine three residues before C4. This suggests that the ligand arrangement observed for EGF11 is not due to its N-terminal character in the hN-1_{11-13} construct.

6.5.4 The EGF4-EGF5 interface

Whereas cbEGF-cbEGF interfaces have been well documented and are understood, interfaces formed by EGF-cbEGF pairs are less studied. The interface of EGF4-5 shows the conserved packing described for the hN-1_{11-12} interface (section 1.4.4). This is for example, a minor β-sheet between the linker residues and the C-terminal residues of EGF4 (H171 and Q177), the water molecule that serves as Ca^{2+} ligand and is held in place by S169, and the conserved aromatic packing residue F170 that packs against V199 in EGF5. V179 packs against H171 and Q177, and residues G169 and G200 form part of the packing region. In section 6.2 it was argued that an additional Ca^{2+} carbonyl ligand at the second strand of the β-sheet may result in an interface with different tilt and twist angles. Here, EGF5 has this ligand too and, with reference to EGF4, shows a larger twist angle (~175°) to what was observed in EGF11-12 and EGF12-13 (~125°) (table 6.1). A structural analysis is presented in Figure 6.24. Using the align function of Pymol, EGF5 and EGF12 of the EGF4-5 and EGF11-12 domain pairs were overlayed, showing a high degree of structural conservation. However, the position of the Ca^{2+} (Box I) is significantly different. In EGF5, it is pulled towards the second strand of the major β-sheet by the backbone carbonyl of S201, whereas the carbonyl of the corresponding residue in EGF12 (E473) is further away (not shown). This causes a bend in the backbone structure at the beginning of EGF5, since the first Ca^{2+} ligand (D178) follows the movement of the Ca^{2+} atom (Box II). This first

ligand has a phi/psi angle of -77°/143° in EGF 5 (D178) and -84°/102° in EGF12. The aromatic packing residue packs from a different angle in the two interfaces (Box III). As explained in the introduction (section 1.4.4.), the aromatic packing residue forms a tight hydrophobic core with many residues from different regions within the same domain. These three aspects may explain the difference in twist angle of around 60°.

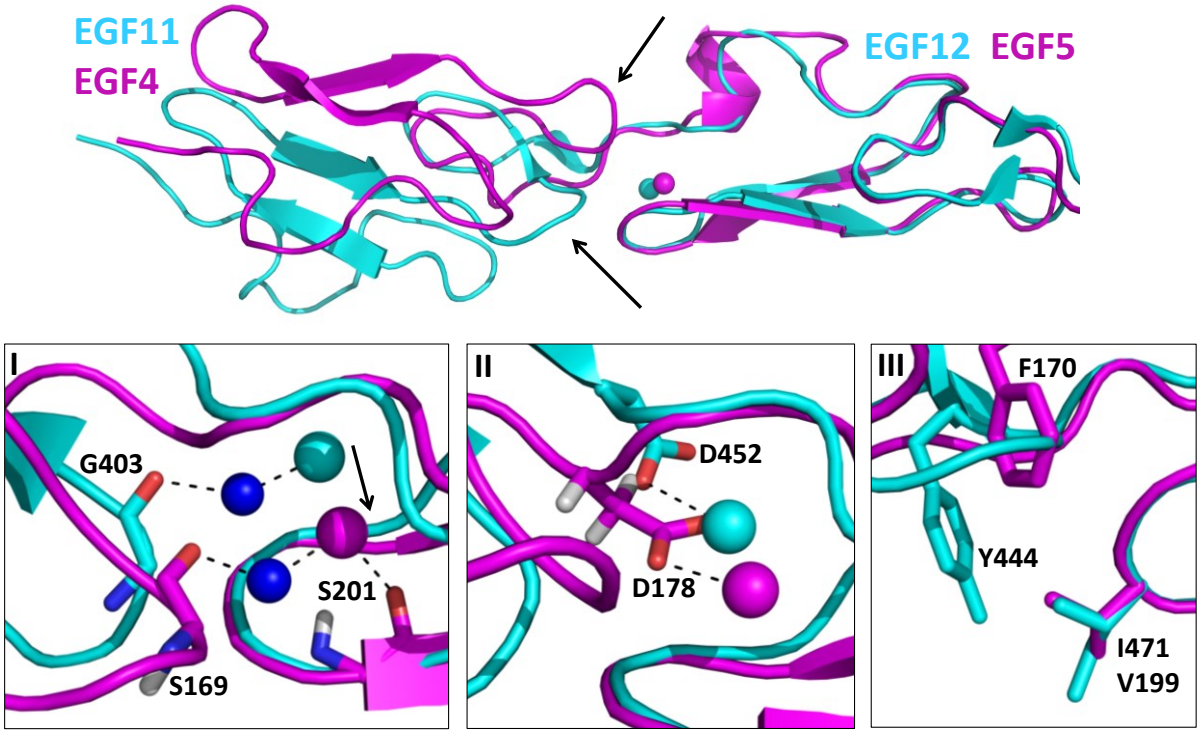


Figure 6.24 Overlay of EGF11-12 (cyan) on EGF4-5 (magenta). Overlaying EGF5 and EGF12 indicates significantly different interdomain twist angles with EGF4 and EGF11, respectively (arrows). Three aspects are highlighted in Box I-III. In EGF5, Ca^{2+} is pulled towards the second half of the β-sheet by the less common carbonyl ligand of S201 (Box I), the arrow indicates the offset direction. This causes a larger distance to the first Ca^{2+} ligand in the sequence which is hence pulled towards Ca^{2+}, causing a bend in the backbone (Box II). The conserved aromatic packing residue (F170) packs to the corresponding hydrophobic residue in a different orientation (V199) (Box III).

Figure 6.25 shows a complementary overlay to the one of Figure 6.24; here EGF11 is overlaid on EGF4. The overlay shows the conservation of backbone structure for EGF4 and EGF11, with a deviation starting in EGF6 at the Ca^{2+} binding region as described in Figure 6.24 (arrow I). However, the packing region in between C5-C6 shows a slight variation too (arrow II). This is due to a change in the corresponding packing regions in the C-terminal domain.

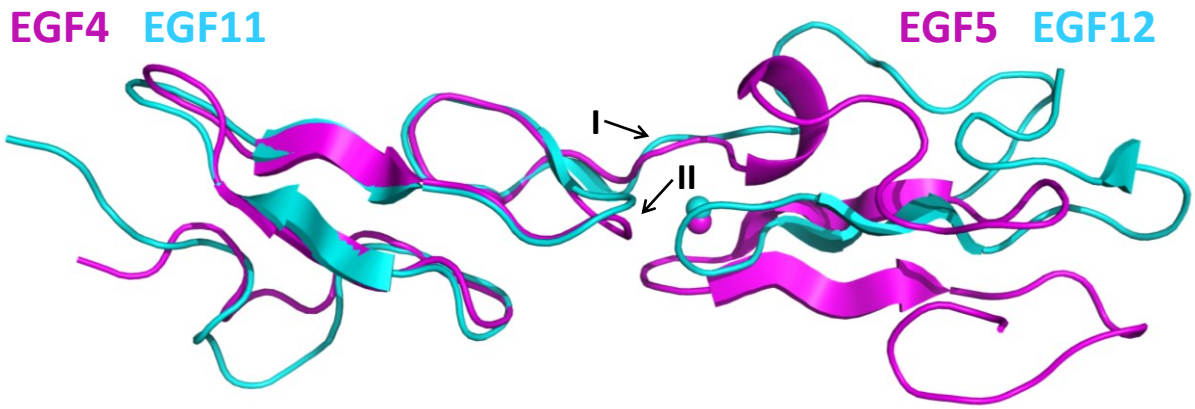


Figure 6.25 Overlay of EGF11-12 (cyan) on EGF4-5 (magenta) on N-terminal domain. Note that EGF11 and EGF4 are structurally conserved, but a difference can be seen for the turn in between C5 and C6 that harbours the consensus aromatic packing residue (arrow II). Arrow (I) indicates the position where EGF12 and EGF5 start to deviate.

6.5.5 The EGF6-EGF7 interface

The EGF6-EGF7 interface shows the conserved aromatic consensus residue in EGF6 (F249) that packs against V276 in EGF7. EGF7 doesn't have a carbonyl Ca^{2+} ligand corresponding to S201 in EGF5 and the interdomain twist angle of EGF6-EGF7 (151.0°) is in between the angles observed for the pairs EGF11-EGF12/EGF12-EGF13 (119.3°/132.0°) and for the EGF4-EGF5 pair (179.7°) (beta-sheet overlay, table 6.1, section 6.4). An overlay of EGF11-EGF12 on EGF6-EGF7 shows that the deviation occurs directly N-terminal of C1 in EGF7/12 (Figure 6.26). Furthermore,

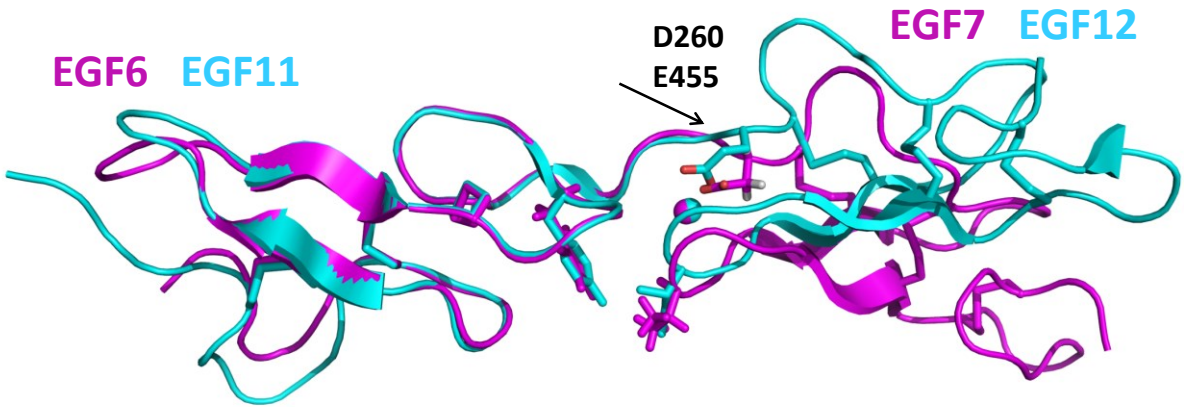


Figure 6.26 Overlay of EGF6-7 (magenta) on EGF11-12 (cyan). Note that the twist angle is slightly different, starting from the residue at position 1 before C1 (arrow); this residue is D260 in EGF7 and E455 in EGF12 and both provide side chain ligands. However, the side of the uncommon aspartic acid is shorter than the common E/Q residue type, causing an alteration of inter-domain tilt and twist angles.

the tilt angle for the EGF6-EGF7 interface (38.6°) is slightly larger than the angles observed for the EGF11-EGF12/EGF12-EGF13/EGF4-EGF5 (9.7°/20.3°/32.8°). Here, the shorter side chain of the uncommon aspartic acid, when compared to the common E/Q residue type, causes the alteration of tilt and twist angles observed.

6.6 Structural analysis of hN-1_{21-23}

6.6.1 The EGF21-EGF22 interface

No structural information was previously available for non Ca^{2+} binding EGF22 (Fig. 6.27).

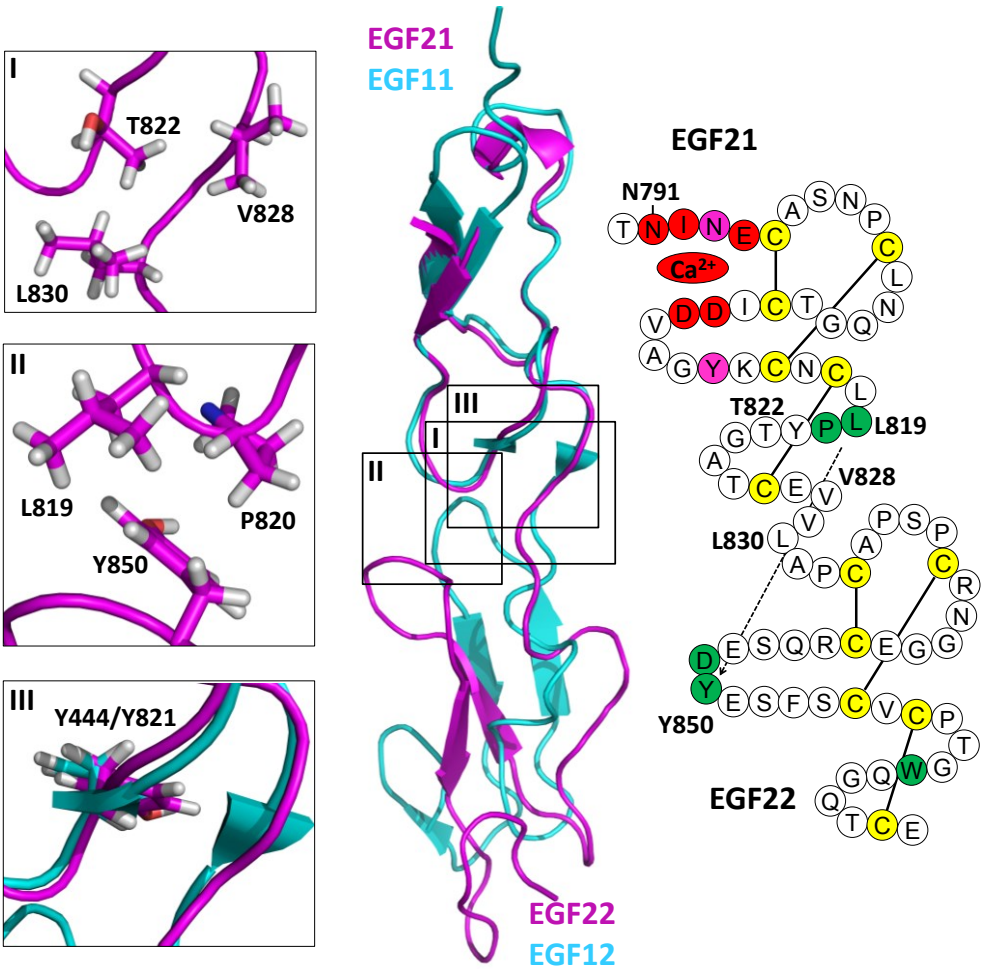


Figure 6.27 Overlay of EGF21-22 (magenta) on EGF11-12 (cyan). Atypical hydrophobic packing is observed between P820 (EGF21) and Y850 (EGF22). The presence of *cis*-proline P820 facilitates this interaction (Box II), but does not distort the positioning of Y821 (Box III). Box I shows the conserved hydrophobic packing of T822 that is in contact with V828 and L830. In the schematic diagram on the right, calcium and ligands are depicted in red, interdomain packing residues in green, cysteines in yellow and conserved residues important for intradomain stability in purple.

The EGF21-EGF22 interface is characterised by a tilt angle close to 0° and a twist angle of around 50° (Figure 6.27) which is significantly different to what had been observed before; for example in the EGF5-6 interface the tilt angle was 90° and the twist angle ~125° (Table 6.1). Many aspects of conventional EGF domain interfaces are present as well, such as the side chain packing of L830 with T822 and V828 (Box I) resulting from the minor β-sheet of T822 and V828 (not shown). Similar to the EGF5-EGF6 interface, no hydrophobic packing was observed between the conserved aromatic packing consensus residue, four residues further than the fifth conserved cysteine, Y821 (EGF21), and residues in EGF22. Instead, a *cis*-proline (P820) together with its neighbouring residue L819 shows close hydrophobic packing with Y850, the second residue at the open end of the large β-sheet in EGF22 (Box II). Interestingly, the *cis*-proline does not distort the backbone and side chain of the conserved aromatic residue Y821 (Box III). The packing of residues at positions 2 and 3 after the fifth conserved cysteine is similar to the packing observed for the EGF5-6 interface, where the main packing residue A208 is situated 2 residues further than the fifth conserved cysteine in EGF5.

Further analysis overlaying EGF12 on EGF22 (Figure 6.28) indicates a different architecture for the major β-sheet in EGF22 (arrow).

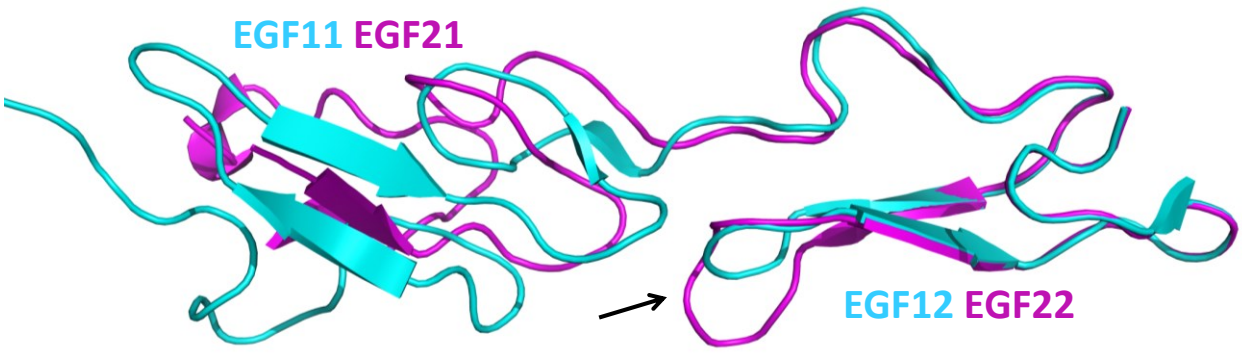


Figure 6.28 Overlay of EGF12 on EGF22. Note the difference in the organisation of the large β-sheet in the C-terminal domain (arrow). The side chains were not shown for simplicity but their position greatly determines the interface angles.

Similarly to EGF6, the major β-sheet has additional residues between C3 and C4 in EGF22. Here, 10 residues are observed, whereas it was 9 residues in EGF6, 10 residues in EGF10 and 8 residues in cbEGF domains of Notch. The longer loop in EGF22 leads to differences in the way that the packing residue interacts with EGF21. The hydrophobic packing residue in this turn, Y850, is located at position 6 after the third conserved cysteine and is the second residue of the beta-turn; this distorts the geometry of the turn and the packing interaction.

6.6.2 The EGF22-EGF23 interface

The EGF22-EGF23 interface showed similar twist and tilt angles as the other EGF-cbEGF interfaces with a twist angle around ~175° and a tilt angle around ~40° (Table 6.1). Overlaying EGF4-5 and EGF22-23 revealed similar interdomain interfaces (Figure 6.29).

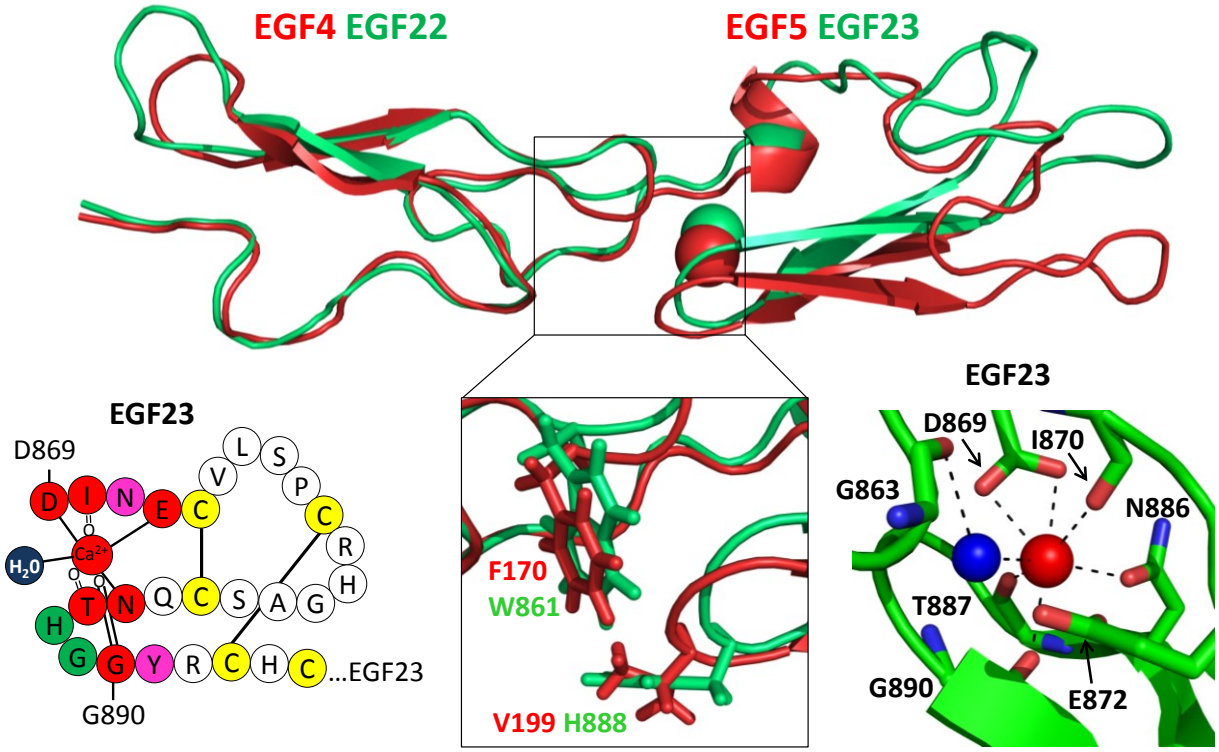


Figure 6.29 Domain interface of EGF4-5 (red) and EGF22-23 (green). Note the similarity of the constructs. This is probably due to a similar Ca^{2+} binding site, since G890 provides a carbonyl Ca^{2+} ligand, similar to what was observed in EGF4-5. Interestingly, the substitution of a tryptophan in EGF22 for a phenylalanine (Box) has little effect on packing (lower middle panel). In the schematic diagram, H_2O is depicted in blue, calcium and ligands in red, cysteines in yellow, inter-domain packing residues in green and conserved residues important for intradomain stability in purple.

A serine in EGF5 and EGF11, at position three before C4, provided a backbone carbonyl Ca^{2+} ligand at this position. This is similar to the cbEGF22-TB4-cbEGF23 interfaces that had a Ca^{2+} binding backbone carbonyl from serine at this position (Lee et al., 2004) or the crystal structure of EMR2, which is another triple cbEGF domain construct (Abbott et al., 2007). In EGF23, this ligand is present as well, but the residue type is a glycine (G890), suggesting that not only serines are able to donate backbone carbonyl ligands. The similarity in the interface between EGF4-5 and EGF22-23 correlates with the positioning of the Ca^{2+} ion and this consistency predicts a similar arrangement for the EGF10-11 interface. Note that in the crystal structure of EGF23, the side chain of E872 is more than 3 Å distance away from the Ca^{2+} ion and this residue is therefore not a Ca^{2+} ligand. The ligands are two side chain residues of D869, one carbonyl of I870, one side chain ligand of E886, and backbone carbonyl ligands of T887 and G890. The seventh ligand is donated by a water molecule.

6.7 Discussion

The analysis of the hN-1_{21-23} and hN-1_{4-7} crystal structures confirmed the structural conservation for cbEGF domains in EGF5, EGF7, and EGF23 when compared to EGF11, EGF12 and EGF13 of the previously published crystal structure of hN-1_{11-13} (Cordle et al., 2008a). Subtle changes in the positioning of the Ca^{2+} ion and specific side chains at the interfaces are responsible for specific interdomain twist and tilt angles.

The interface of EGF12-EGF13 is slightly more curved when compared to EGF11-EGF12 because of the displacement of the packing residues Y482 and I509 by the amide group of the side chain of the first Ca^{2+} ligand in EGF13, N490. Most residues belonging to the EGF backbone show very distinct torsion angles depending on their position in the domain. The difference in interdomain tilt and twist angles observed for the EGF11-EGF12 and EGF12-13 domain pair is reflected over a wide stretch of amino acids with subtle individual changes and cannot be deferred from the torsion angles.

Furthermore, Ca^{2+} binding in cbEGF domains can be divided into two classes. EGF5, EGF11 and EGF23 have a serine/glycine residue at position three before C4 that donates a backbone Ca^{2+} ligand. These cbEGF domains have tilt angles of around ~40° and twist angles of around ~175° to their preceding domain. Compared to that, the corresponding residue in EGF7, EGF12 and EGF13 does not donate a carbonyl backbone Ca^{2+} ligand. The twist and tilt angles with their preceding domains are hence ~125° and 15° for EGF12 and EGF13 and around 150°/40° in EGF7. It is interesting to speculate why this residue donates a Ca^{2+} ligand only in certain cases. In EGF7, this residue is a threonine, whereas it is a glutamic acid in EGF12 and EGF13. It is a serine in EGF5 and EGF11, and a glycine in EGF23. The study of the cbEGF22-TB4-cbEGF23 construct suggested that this residue needs to be a serine or a threonine to provide the correct hydrogen bonds to neighbouring domains with their hydroxyl group (Lee et al., 2004), but the example of EGF23

shows that it can be as well a glycine. Here, the overlay of EGF11-EGF12 on EGF4-EGF5 (Figure 6.30) shows that the main carbonyl points in the same direction when compared to S201. It is therefore not quite clear why only Serine can bind Ca^{2+}.

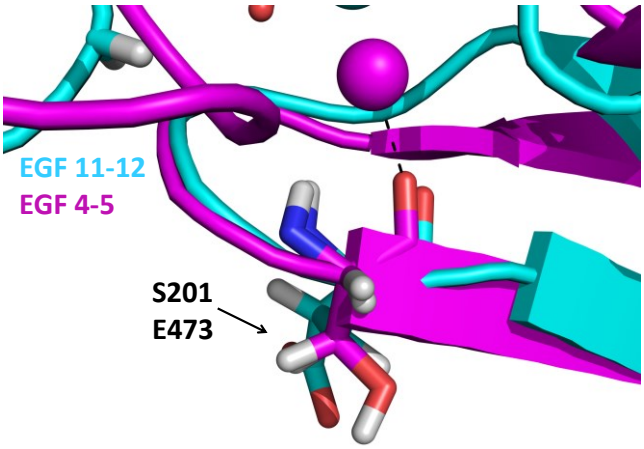


Figure 6.30 Ca^{2+} ligand properties of S201 in EGF5 (magenta) and E473 in EGF12 (cyan). The overlay shows that the carbonyl and NH of the main chain as well as the side chains of S201 and E473 have a similar orientation within the turn of the β-sheet.

EGF7 is different since it has an uncommon aspartic acid in position one before the first conserved cysteine; aspartic acid contains a shorter side chain than Q/E residues which are usually found at this position and results in a more twisted angle when compared to the EGF11-EGF12 and EGF12-EGF13 interfaces. This is possibly similar on the EGF8-EGF9 interface since EGF9 has an aspartic acid at this position too. This analysis shows that it might be possible to predict domain orientations of a cbEGF with its preceding domain based on the position and amino acid type of the Ca^{2+} ligands, since several conserved aspects of cbEGF domains do not allow the full range of tilt and twist angles. For example, the residue at position three after the fifth conserved cysteine provides a hydrogen bond to a water molecule that serves as a Ca^{2+} ligand in all cbEGF domains observed so far. This therefore restricts the distance of this residue to the Ca^{2+} ion. The packing residues are then possibly arranged around this position. However, more structural information is necessary to test these predictions and look for exceptions.

The interfaces formed by cbEGF-EGF domains, as here demonstrated by the EGF5-6 and EGF21-EGF22 interfaces vary more. This variety is in line with the predictions that EGF domains could cause an alteration of the overall structure of Notch (Hambleton et al., 2004). The interface of EGF5-6 shows a bend of around 80-90°. It is not yet clear how rigid this interface is in solution, but the NMR NOE measurements indicate contacts between the crucial A208 and V239 side chains, demonstrating that this interface is not an artefact of crystal contacts. The interface of EGF21-EGF22 shows a tight packing with an aromatic residue (Y850) at the C-terminal domain that packs against a *cis*-proline (P820) one position before the conserved aromatic packing consensus residue (Y821) at the C-terminus of EGF21. The conserved aromatic residue (Y821) did not show any contacts to EGF22. The fact that constructs containing EGF21-22 and EGF5-6 crystallised, suggest that the prediction of these sites to be flexible is probably not right. However, EGF9-10, the third interface under observation did not crystallise, similar to the hN-$1_{11\text{-}13}$12DG mutant in which Ca^{2+} binding was abrogated in EGF12 which may suggest it is more flexible. This indirect evidence for flexibility will be addressed in chapter 7 and 8.

The structural information obtained in this chapter is helpful, but further biological data is needed to understand the structure-function relationship. For example, an inhibitory effect of EGF10 on Notch binding to Delta-like 1 protein was measured in FACS analysis (Cordle et al., 2008b) which was dependent on Ca^{2+} binding in EGF11. Whereas EGF11-EGF12 is rigid, this analysis suggests a slightly larger twist and tilt angle for the EGF10-EGF11 interface that could potentially cause steric hindrance of the Delta-like 1 ligand. More ligand-binding and activation data are necessary to complement the structural data shown here.

7 TALOS+ analysis

7.1 Introduction to TALOS+ and dynamics

In the previous chapter, the analysis of high resolution structures obtained by X-ray crystallography for the hN-1_{4-7} and hN-1_{21-23} constructs was presented. In this chapter, NMR chemical shifts are analysed with TALOS+, a program that is used to give complementary information. Furthermore, structural information will be presented for the hN-1_{8-11} construct that has not crystallised.

Chemical shifts are highly dependent on local structure. Recently, a method has been developed to predict protein backbone torsion angles ϕ/ψ of residue i using HN/Hα/CO/Cα/Cβ/N chemical shifts of three sequential residues, i-1, i and i+1 by matching these to a database of 200 proteins (TALOS+) (Shen et al., 2009). Using a neural network filter, probabilities of the residue in question to be part of either alpha, beta or loop regions are made based on the position of the 10 best matches of the database query and their position in the Ramachandran plot. For each residue, TALOS+ assigns the attribute "good" to the prediction, if all 10 database matches are in a consistent region of the Ramachandran plot, the predicted order parameter S^2 value is > 0.5, and the probability is significantly higher for the residue to adopt one of the three states alpha, beta or loop (ANN classification system). The attribute "ambiguous" is added, if these requirements are not fulfilled. In a test data set, consistent torsion angles with the attribute "good" were predicted for an average of 88% residues; 2.5% of these were wrong predictions. Excluding the 2.5% wrong predictions, an accuracy of ±13° was found for ϕ and ψ angles. Predictions were wrong for loop regions and regions which were poorly defined in crystal structures. Chemical shifts are also used in *de novo* protein structure generation using the standard ROSETTA Monte Carlo assembly method (Shen et al., 2008, Bradley et al., 2005). However, the authors reported certain difficulties in complex proteins such as proteins with prosthetic groups and little secondary structure. EGF-like domain constructs show little secondary structure, but are predominantly folded by disulphide bond formation and

Ca^{2+} binding. Therefore, it is not clear whether the TALOS+ and ROSETTA approaches are able to reveal reliable structural information for EGF-like domains.

In this chapter, TALOS+ is first tested on the hN-1_{11-13} (Figure 7.1) and hN-1_{4-7} (Figure 7.2 left) constructs for which crystal structures are available. NMR chemical shifts are recorded at pH 6.1 and under Ca^{2+} saturation (30 mM Ca^{2+}).

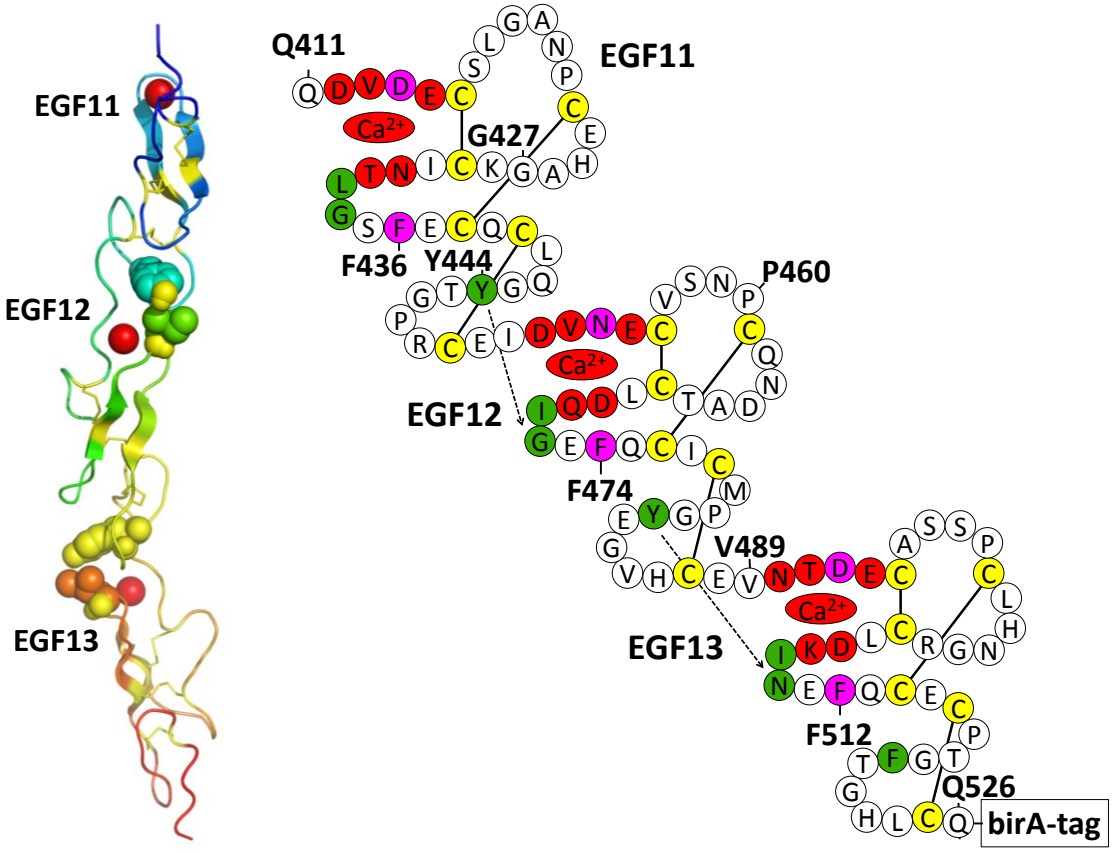


Figure 7.1 Crystal structure of hN-1_{11-13} (left) and the corresponding schematic diagram (right). Note that the birA-tag in this construct (Cordle et al., 2008a) was present in the sample used for crystallography, but not for the sample used for chemical shift assignments (Muranyi et al., 2004). The addition of the birA-tag is thought to stabilise EGF13. Interdomain packing residues are indicated by spheres, Ca^{2+} is shown in red. In the schematic diagram, calcium and ligands are depicted in red, interdomain packing residues in green, cysteines in yellow and conserved residues for intradomain stability in purple.

TALOS+ will then be used to predict structural properties of the hN-1_{8-11} construct (Figure 7.2 right). This analysis will provide information regarding the reliability of using TALOS+ restraints in structure determination of EGF domain constructs by NMR.

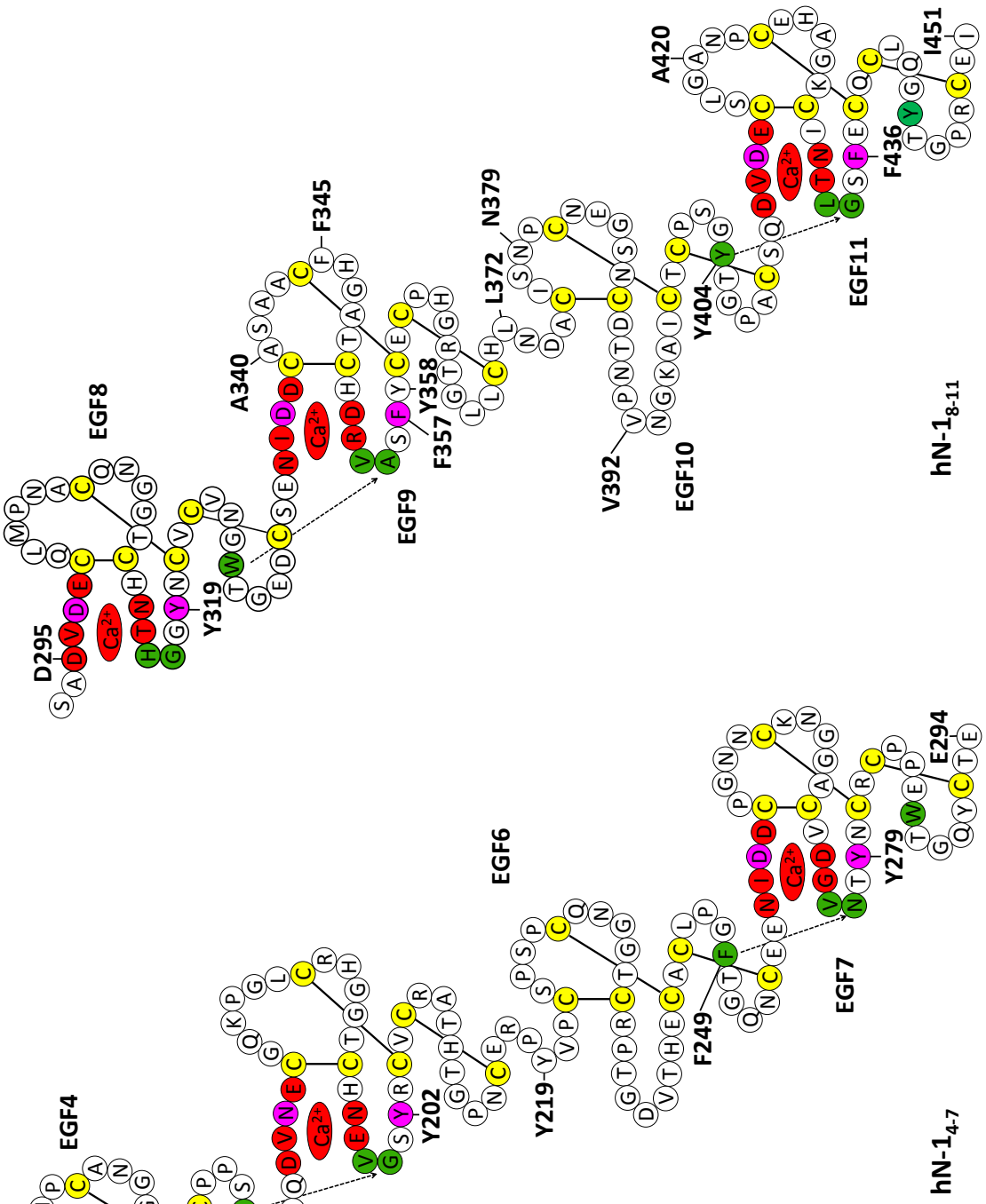


Figure 7.2 Schematic diagrams of hN-1$_{4-7}$ (left) and hN-1$_{8-11}$ (right) for which a TALOS+ analysis is presented. In the schematic diagram, calcium and ligands are depicted in red, interdomain packing residues in green, cysteines in yellow, and conserved residues for intradomain stability in purple.

7.2 TALOS+ torsion angle predictions

7.2.1 Predictions for individual domains in hN-1_{11-13}

The phi and psi torsion angles of the individual domains EGF11, EGF12 and EGF13, derived from the crystal structure are plotted in Figure 7.3 as red triangles. Predictions for ϕ and ψ were based on the chemical shifts on hN-1_{11-13} (Muranyi et al., 2004) (black circles in Figure 7.3). For many residues, the predictions based on chemical shifts are very close to the torsion angles derived from the crystal structure.

In the construct of hN-1_{11-13} used for crystallography, a stabilising birA-tag was added at the C-terminus, when compared to the construct for which chemical shifts were obtained, this explains the worse quality of predictions for EGF13. In general, this analysis shows that most of the predictions are correct, with the exceptions of several glycines, the loop region and residues with missing assignments. Interestingly, the glycine at the turn, four residues before C6 showed the same deviation from the correct phi angle in both EGF11 and EGF12, whereas the assignments were missing for EGF13.

Predictions of torsion angles within the loop are different to the crystal structure most likely due to the flexible nature of this region (Hambleton et al., 2004). The corresponding region was also poorly defined when comparing the different crystal structures of hN-1_{4-7} (section 6.5.1), suggesting that the determination of the torsion angles in this region by crystallography is imprecise and that TALOS+ might be able to provide better information regarding these sites.

As outlined in the previous chapter, subtle changes in torsion angles at key residues can have a long-range effect on inter-domain tilt and twist angles. Therefore, an analysis of the precision and accuracy of TALOS+ is necessary to understand when and how predictions fail; torsion angle restraints with little accuracy can then be excluded from further analysis.

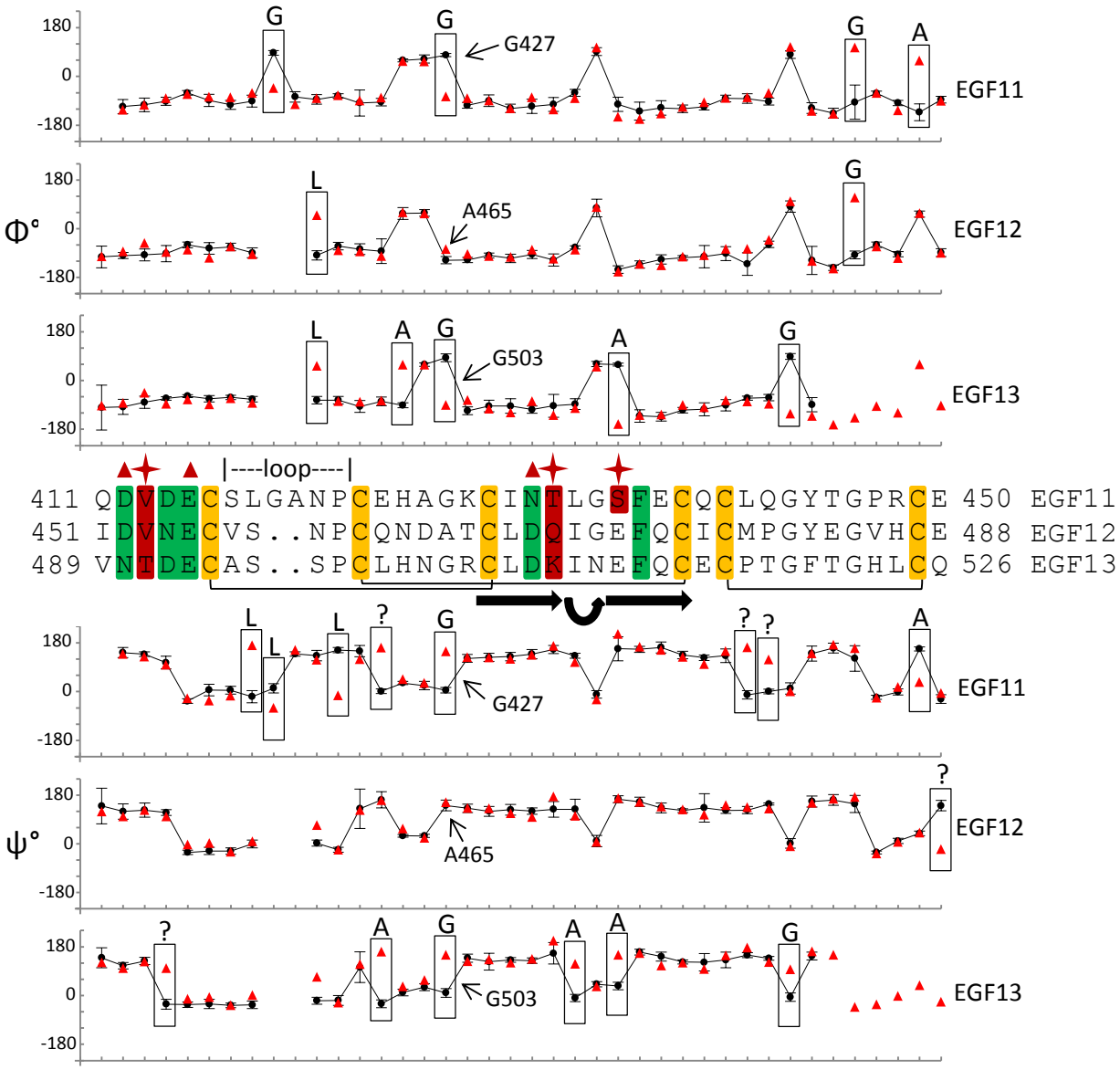


Figure 7.3 Prediction of torsion angles based on chemical shifts of hN-$1_{11\text{-}13}$. Black circles indicate the angle predicted from the chemical shifts, the error bars show the standard deviation based on the difference of the 10 best matches of the database query. Red triangles indicate torsion angles derived from the crystal structure. The schematic diagram in the middle shows the alignment of individual domains. Consensus Ca^{2+} binding residues are shown in green, cysteines in yellow. Other Ca^{2+} ligands are indicated in red, with a red triangle for side chain ligands and a red star for carbonyl backbone ligands. The position of the loop in between C1 and C2 is indicated as well as the position of the β-sheet (arrows). Note that overall good agreement is observed. Among the residues with very poor predictions (black boxes) are residues of the loop region between C1 and C2 (L) and glycines (G). Residues for which less than 17 shifts were assigned are indicated with (A). All major violations of phi angles match one of these categories, whereas the psi angle of five residues did not (question mark). Most phi/psi angles are very similar among the EGF domains. The very first residues of EGF11 are different as EGF11 is the N-terminal domain. Two conserved glycines at the C-terminus of EGF13 also show different angles. Note that the N-terminal residue in EGF11 was not present in the crystal structure.

7.2.2 Accuracy of torsion angle predictions in hN-$1_{11\text{-}13}$

Here, 109 torsion angle predictions from D412 to F520 were compared to the crystal structure, 79 of which were attributed as “good” and 30 as “ambiguous”. Of the 79 torsion angles predicted as “good”, only 62 had both the ϕ and ψ torsion angles within 30° of the crystal structure. Figure 7.4 shows the TALOS+ output as Ramachandran plots of four representative residues with the crystal structure angles as a blue square (white arrow).

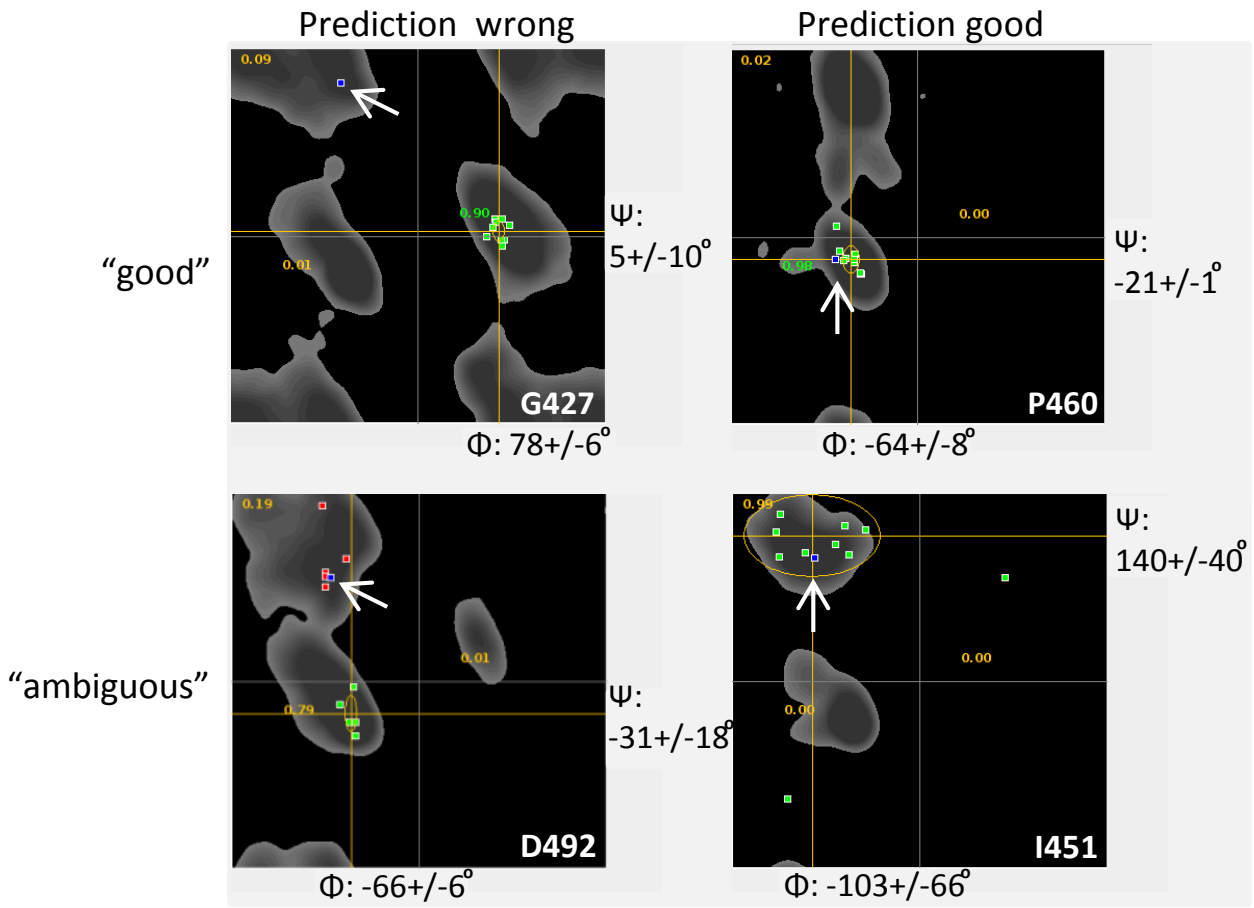


Figure 7.4 Ramachandran plots of 4 representative residues based on a TALOS+ search. Depicted in grey are the positions most populated in the database of a certain residue type. Predictions (orange circle and intersection) were either attributed with “good” (top panels) or “ambiguous” (bottom panels), based on either consistent or inconsistent distribution of the 10 best matches of the database query (green and red squares). Predictions were either correct as for P460 and I451 (<30° deviation of the crystal structure, right) or wrong as shown for G427 and D492 (left). The white arrow indicates the position of the ϕ/ψ torsion angles derived from the crystal structure (blue squares).

The 10 matches obtained from the database query are shown in red and green squares. G427 and P460 (top) were attributed as “good”. However, the predictions were only correct for P460. The bottom panels show predictions which were attributed as “ambiguous”. In D492 (lower left panel), 5 matches were close to the actual angles in the β-sheet region, whereas 5 residues were found in the α-helical region. As the matches for the β-sheet region were closer together than the helical ones, the program wrongly concluded the angle to be in the β-sheet region. Since this residue is a Ca^{2+} binding ligand, it is possible that the addition of chemical shifts from EGF-like domain constructs would improve this prediction. For I451, 8 matches were spread around the actual angles, and 2 residues were in atypical regions. Although the predictions for I451 were attributed as “ambiguous”, the angles of 8/10 predictions were close to the crystal structure. 14 of the 30 torsion angles with the attribute “ambiguous” were as well within 30° of the crystal structure.

Many glycines show wrongly predicted angles including the conserved glycine three residues before C6 in EGF11 and EG12, whereas chemical shifts were missing for this position in EGF13. Predictions for 6 out of 10 glycines were wrong. This is most likely due to the unique character of glycines that show more conformational freedom than any other amino acid. Similarly no C_β shift is possible for this residue type, so predictions are based on a smaller number of shifts. G427 (EGF11) and G503 (EGF13), for example have wrong predictions, whereas the predictions for A465 (EGF12) are correct – these three residues are at homologous position in the sequence. This finding is in line with the test-run provided by Ad Bax and colleagues where glycines had the highest percentage of prediction failures (Shen et al., 2009). Most cysteines, most residues involved in packing, and side chain Ca^{2+} ligands were predicted well. However, most angles derived for a residue that provides a backbone carbonyl Ca^{2+} ligand are not correct, including S435, V453, Q470, T491 and K508, whereas good agreement was found for V413 and T432. This analysis shows that chemical shifts of cbEGF domains can be used to predict phi/psi

torsion angles in these domains, if certain difficulties for glycines and for Ca^{2+} backbone ligands are taken into account. The prediction is less accurate for residues with missing assignments. The lack of confidence for the prediction of glycine torsion angles is unfortunate since glycines are located at key positions in EGF domains where they enable the formation of turns. Table 7.1 summarises details of this analysis, showing all residues that have a difference of 30° or more in at least one torsion angle.

Table 7.1 Residues with predicted torsion angles of hN-$1_{11\text{-}13}$ with 30° or more deviation from the crystal structure.

residue	phi	psi	shifts	database	prediction	explanation
C416	ok	off	18	imprecise	good	loop region
L418	off	off	17	3 ok, 7 off	ambiguous	loop region
G419	off	off	17	inaccurate	good	glyine, loop region
A420	off	ok	16	imprecise	good	loop region
P422	ok	off	15	2 ok, 8 off	ambiguous	loop region
C423	ok	off	16	imprecise	ambiguous	loop region
E424	ok	off	18	5 ok, 5 off	ambiguous	imprecise database query
G427	off	off	17	inaccurate	good	glycine
S435	off	off	17	imprecise	ambiguous	backbone ligand
L441	ok	off	18	inaccurate	good	imprecise database query
Q442	ok	off	17	2 ok, 8 off	ambiguous	imprecise database query
G446	off	off	14	inaccurate	ambiguous	glycine, missing assignments
C449	off	off	15	inaccurate	good	missing assignments
V453	off	ok	18	imprecise	good	backbone ligand
N459	off	off	15	2 ok, 6 off	ambiguous	loop region
A465	off	ok	15	9 ok, 1 off	ambiguous	missing assignments
Q470	ok	off	18	imprecise	good	backbone ligand
M479	off	ok	15	imprecise	ambiguous	missing assignments
G484	off	ok	17	inaccurate	good	glycine
E488	ok	off	17	inaccurate	ambiguous	unknown
T491	off	ok	16	imprecise	good	backbone ligand
D492	ok	off	16	5 ok, 5 off	ambiguous	missing assignments
S496	ok	off	17	imprecise	good	loop region
S497	off	off	15	inaccurate	ambiguous	loop region
L500	ok	off	14	inaccurate	good	missing assignments
H501	off	ok	15	inaccurate	ambiguous	missing assignments
G503	off	off	15	inaccurate	ambiguous	glycine, missing assignments
R504	off	ok	14	imprecise	good	missing assignments
D507	off	ok	16	imprecise	good	missing assignments, side chain ligand
K508	off	off	13	imprecise	ambiguous	backbone ligand, missing assignments
I509	ok	off	11	inaccurate	good	missing assignments
E511	ok	off	12	inaccurate	ambiguous	missing assignments
Q513	ok	off	15	imprecise	good	missing assignments
G519	off	off	16	inaccurate	good	glycine, missing assignments
F520	off	ok	10	inaccurate	good	missing assignments

The predictions were either attributed with “good” or “ambiquous” (Shen et al., 2009) by TALOS+. The number of chemical shifts is indicated; the maximum is 18 shifts if all assignments were reported. When significantly fewer shifts were reported, the predictions were inaccurate as in the case of F520 for which only 10 shifts were available. The database column indicates the distribution of the 10 matches of the database query relative to the correct values. The database search was assigned “inaccurate” if matches were close together but far away from the crystal structure value. The attribute “imprecise” was added when values were spread around the crystal structure values. Sometimes, matches were clustered into a group that was close to the crystal structure value (ok, database column) and another group that was not close at all to the crystal structure value (off, database column). The fact that backbone carbonyl ligands are mostly giving bad predictions is probably linked to the effect of the Ca^{2+} ligand on the chemical environment of the chemical shifts nearby and not to a specific torsion angle. This could possibly be improved by taking the assignments of hN-$1_{4\text{-}7}$ and hN-$1_{11\text{-}13}$ into the chemical shift database.

7.2.3 Predictions for individual domains in hN-$1_{4\text{-}7}$

Figure 7.5 shows the overlay of EGF domain specific torsion angles in the hN-$1_{4\text{-}7}$ structure. Similar to hN-$1_{11\text{-}13}$, the most variation of angles is observed between C1 and C2, with EGF6 having a distinctly different psi angle for C1, probably due to the position of a *cis*-proline two positions later. The boxes highlight where differences occur.

Figure 7.6 shows the TALOS+ analysis of EGF4 as compared to the torsion angles from the crystal structure. Predictions were not good for the loop residue N147, glycine G172 and for A150. The remaining predictions were within or close to the errors obtained. The error for the prediction of C149 was high. Residues around A160 showed large errors even though the assignments for this region were complete. This is consistent with the lack of electron density in this region. This could possibly be overcome by the addition of EGF3 to the construct to stabilise

this region. However, the torsion angles predicted may help modelling this region in combination with additional data from spin-spin couplings or NOEs.

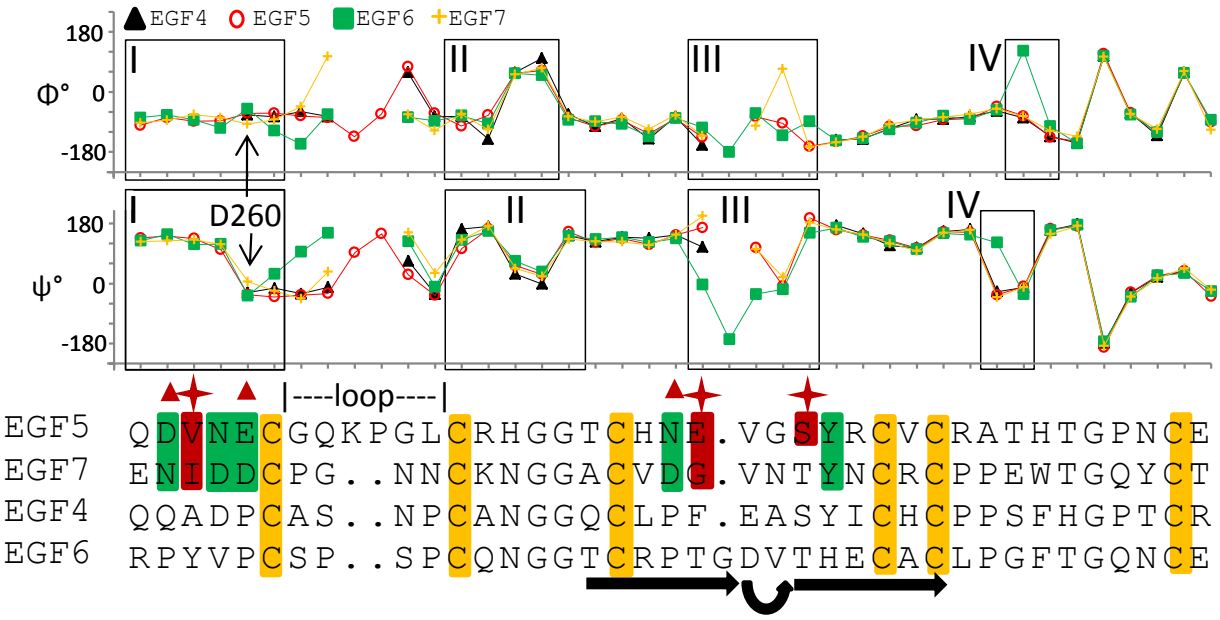


Figure 7.5 Phi/Psi torsion angles from the crystal structure of hN-1$_{4-7}$. Overall a consistent pattern of the EGF fold is observed. Torsion angles vary within the loop region of C1 and C2. Boxes highlight other regions were structures deviate (I-IV). Little difference is seen at the beginning of the domains (Box I), however subtle differences in this region are crucial to adopt specific inter-domain angles. This was demonstrated for EGF6-EGF7 (section 6.5.5), where the less common residue type aspartic acid (D260, EGF7) caused an altered interface. Box II shows subtle changes. In Box III, the variations of the turn in the major beta-sheets indicate large differences. In Box IV, EGF6 has distinct torsion angles, whereas the other EGF domains have similar torsion angles. In the sequence alignment, consensus Ca^{2+} binding residues are highlighted in green, cysteines in yellow, other Ca^{2+} ligands in red, with a red triangle for side chain ligands and a red star for carbonyl backbone ligands. The position of the β-sheet is indicated (arrows).

Predictions at the C-terminus of the domain were accurate and show the clear difference observed between EGF4 and EGF6 (Box IV in Figure 7.5). Figure 7.7 shows the prediction for EGF5. Similar to EGF4, L183 of the loop region was not predicted well nor was G212 at a homologous position to G172. Assignments were missing for E216. Similar to EGF4, the psi angle of the first residue after C2 was not predicted well (R190). The reason for the failure of predictions for C189, R207 and C215 are unclear and may show a region with differences in solution and X-ray structure.

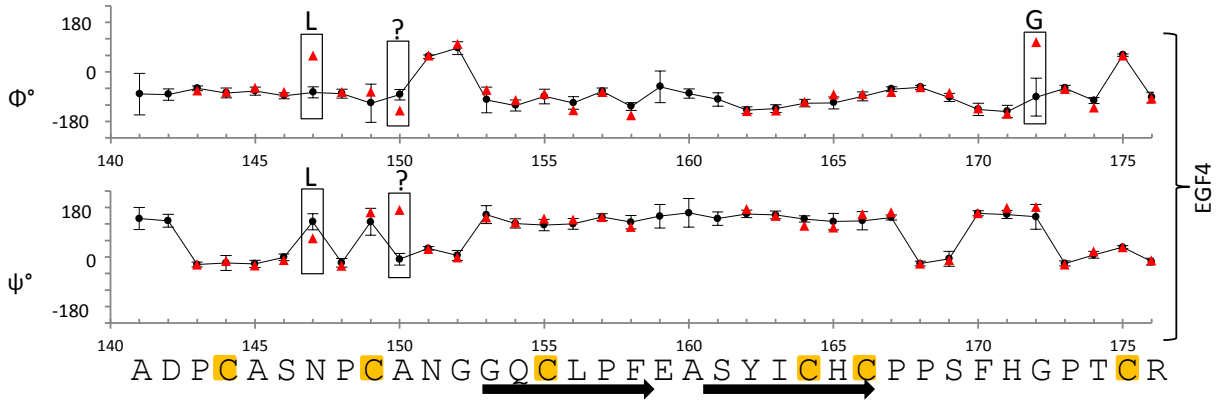


Figure 7.6 TALOS+ analysis of EGF4. Most torsion angles are well predicted. Exceptions are indicated in boxes, G for glycines, L for loop region, and question mark where prediction failed without any obvious reason. The region before C1 is not well defined. Cysteines are highlighted in yellow, black arrows indicate the beta-sheet, red triangles indicate the torsion angles derived from the crystal structure and the black circles indicate the TALOS+ predictions.

Interestingly, the first residues within EGF5 were predicted well; this is an interesting result given that these residues are sensitive to interdomain angles.

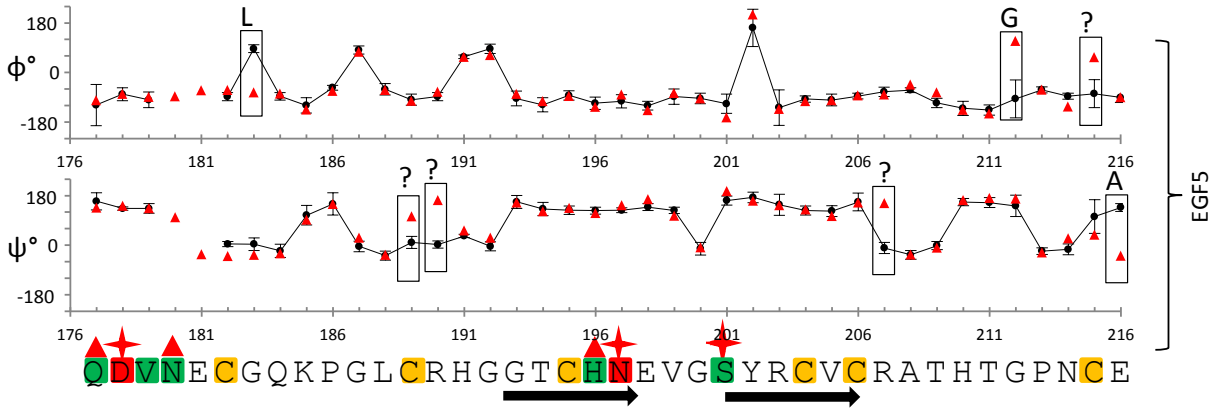


Figure 7.7 TALOS+ analysis of EGF5. Most torsion angles are well predicted. Exceptions are indicated in boxes (G) for glycines, (L) for loop region, (A) for missing assignments and (?) where prediction failed without any obvious reason. Assignments were missing around the region of C1. Consensus Ca^{2+} binding residues are highlighted in green, cysteines in yellow, other Ca^{2+} ligands in red, with a red triangle for side chain ligands and a red star for carbonyl backbone ligands. The position of the β-sheet is indicated (arrows).

The prediction of EGF6 shows wrong phi angle predictions for G237 and G251 (Figure 7.8). The prediction of the psi angle of P221 was wrong as well, most likely due to a lack of assignments. P226 was wrongly predicted, this residue is a *cis*-proline and the prediction most likely is wrong because of its position in the loop region between C1 and C2 or a lack of *cis*-

proline in the database. Key residues within the different boxes were predicted well. Interestingly, EGF6 has a different torsion angle at position two and three after C5 (P247 and G248, Figure 7.5 and 7.8) and to a minor extent at F249, when compared to EGF4, EGF5 and EGF7, these are predicted well by TALOS+. This region is very important for the domain interaction between EGF6 and EGF7.

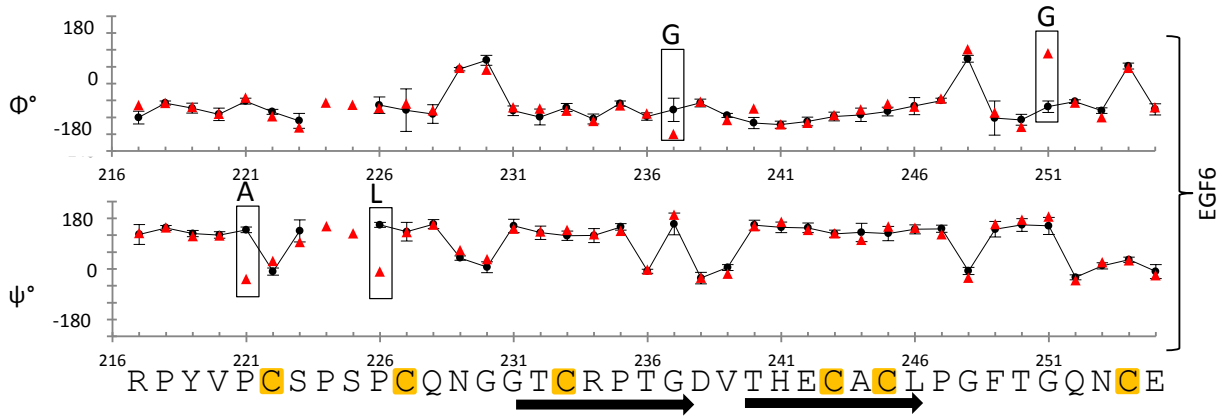


Figure 7.8 The phi/psi torsion angle predictions of EGF6. Similar to previous domains, two glycines (G) showed wrong predictions. P221 was predicted wrong, most likely because of missing assignments (A) and P226 was within the loop region (L). Black arrows indicate the beta-sheet, red triangles indicate the torsion angles derived from the crystal structure and the black circles indicate the TALOS+ predictions.

Figure 7.9 shows relevant interdomain contacts. The conserved aromatic packing residue F249 packs with V276.

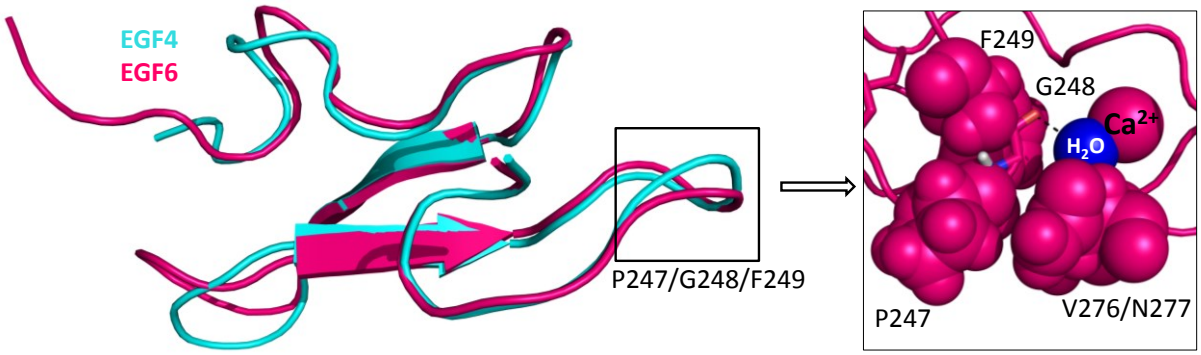


Figure 7.9 Structural differences between EGF4 (cyan) and EGF6 (magenta). Three residues of EGF6, P247, G248 and F249 are interacting with EGF7 (right hand panel). F249 interacts with V276 (on the back of the right panel). P247 interacts with N277 and G248 shows the conserved hydrogen bond to the H_2O Ca^{2+} ligand.

P247 interacts with N277 and G248 is involved in holding the H_2O Ca^{2+} ligand in place. This is an interesting result showing that chemical shift predictions can in certain cases reveal information about the interfaces formed. In EGF7, the prediction of G289 was wrong, consistent with previous domains (Figure 7.10). Further unsuccessful predictions were N277, the residue that packs with EGF6, P262 of the loop region, and T293, the second to last residue of the sequence.

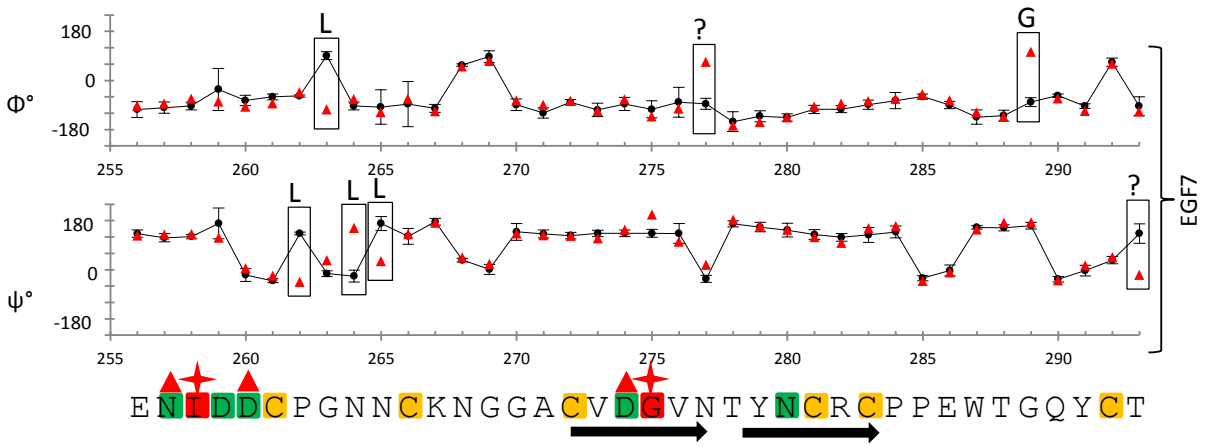


Figure 7.10 Phi/Psi torsion angle predictions of EGF7. Similarly, to previous domains, the prediction of G289 (G) was wrong as were predictions for the loop region (L). Predictions for N277 and T293 as indicated by a question mark (?) failed without any obvious reasons. Other predictions agreed quite well. Black arrows indicate the beta-sheet, red triangles indicate the torsion angles derived from the crystal structure and the black circles indicate the TALOS+ predictions. Cysteines are highlighted in yellow, Ca^{2+} binding consensus residues in green, carbonyl backbone Ca^{2+} ligands in red; the red triangle indicates Ca^{2+} side chain ligands, and the red asterisk Ca^{2+} backbone ligands.

The predictions of the very first residues of EGF7 were within the error range; however it was outlined in section 6.5.5 that slightly different torsion angles around D260 lead to an altered interface because of the shorter side chain of D when compared to other domains containing an E or Q at this position. The deviation of angles for D260 (Figure 7.5, arrow) is quite small around 30° and comparable to the error involved in the torsion angle prediction. Predictions after the second conserved cysteine were more precise, but not all predictions within the major β-sheet did agree well; for example N277 at the turn did not agree well. The predictions for residues at the C-terminus of EGF7 were good. Phi angles of the glycine residue at position 3 before C6 are consistently incorrectly predicted by a similar amount. This could possibly be adjusted in the

TALOS+ by adding these assignments to the database. However, it cannot be excluded that this difference comes from the way the X-ray structure refinement is done or a difference between solution and X-ray structure. The predictions at the C-terminus, when excluding this glycine are good and can be used to indicate whether or not an EGF domain falls within the category of EGF4, EGF5 and EGF7 or if they are more like EGF6. The predictions around the turn in the beta-sheet are often not correct, such as in EGF7. Predictions at the N-terminus of the construct are probably reasonable, but not accurate enough for structure determinations.

7.2.4 Accuracy of torsion angle predictions in hN-$1_{4\text{-}7}$

Torsion angles were predicted for 149 residues from A141 to T293. Residues A141, D142, E159, A160 and S161 weren't present in the crystal structure. From the remaining 144 predicted torsion angles, 44 were assigned with the attribute "ambiguous", and 100 with the attribute "good". Table 7.2 summarises residues with a deviation of more than 30° for one or both torsion angles. Similarly to hN-$1_{11\text{-}13}$, predictions were reasonable for cysteines and side chain Ca^{2+} ligands. 2 out of 5 backbone ligands were wrongly predicted. From the 100 residues with the attribute "good", only 15 showed a deviation of more than 30° for one or both torsion angles. From these 15 residues, 6 were glycines with three of them having deviations smaller than 40°, and three of them deviations larger than 40°. From the other 9 residues, N277, N264, T240 and C189 showed deviations larger than 50°, the remaining residues smaller than 50°. Whereas N264 is in the potential flexible loop of EGF7, C189 is located at the end of the loop in EGF5, and only 9 out of 18 possible chemical shifts were available for this residue. For N277, all 18 shifts were available, but the prediction of its phi angle was wrong by 154° and its psi angle was wrong by 52°. It needs to be noted that N277 is the second residue of the turn in the beta sheet of EGF7, a position usually adopted by a glycine. In the crystal structure, the side chain of this residue points towards the hydrophobic packing site. Whether this conformation is true in solution, is not known. T240 is also the residue before the turn in the beta sheet of EGF6.

Table 7.2 Residues with predicted torsion angles of hN-$1_{4\text{-}7}$ with 30° or more deviation from the crystal structure.

residue	phi	psi	assigned	database	predict	explanation
N147	off	off	15	inaccurate	ambi.	loop region
C149	off	off	16	unknown	ambi.	loop region
A150	off	off	18	4 ok, 6 off	ambi.	imprecise database query
G153	off	ok	16	imprecise	ambi.	glycine, only slightly off
F158	off	ok	16	imprecise	good	undefined in crystal structure
G172	off	off	14	inaccurate	ambi.	glycine
C182	ok	off	9	unknown	good	missing assignments, loop region
G183	off	off	15	inaccurate	ambi.	glycine, loop region
C189	ok	off	17	inaccurate	ambi.	loop region
R190	ok	off	17	4 ok, 6 off	ambi.	imprecise database query
G192	ok	off	16	imprecise	good	glycine, only slightly off
S201	off	off	17	inaccurate	ambi.	Ca^{2+} backbone ligand
Y202	off	ok	17	imprecise	ambi.	unknown
R207	ok	off	18	1 ok, 9 off	ambi.	imprecise database query
T209	off	ok	18	imprecise	good	unknown, just slightly off
G212	off	ok	15	inaccurate	ambi.	glycine
N214	off	off	15	imprecise	good	missing assignments, only slightly off
C215	off	off	17	unknown	ambi.	unknown
E216	ok	off	15	1 ok, 9 off	ambi.	imprecise database query
R217	off	ok	14	imprecise	ambi.	missing assignments
P221	ok	off	12	inaccurate	ambi.	missing assignments
S223	ok	off	10	inaccurate	ambi.	missing assignments, loop region
P226	ok	off	10	inaccurate	ambi.	missing assignments, loop region
G230	off	ok	16	imprecise	good	glycine, only slightly off
T232	off	ok	17	imprecise	good	unknown, only slightly off
G237	off	off	17	inaccurate	ambi.	glycine
T240	off	ok	18	imprecise	good	unknown, only slightly off
G248	off	off	15	imprecise	good	glycine, only slightly off
G251	off	off	17	inaccurate	good	glycine
D259	off	off	18	5 ok, 5 off	ambi.	imprecise database query
P262	ok	off	15	2 ok, 8 off	ambi.	imprecise database query
G263	off	off	15	inaccurate	ambi.	glycine
N264	ok	off	17	inaccurate	ambi.	loop region
N265	ok	off	18	inaccurate	ambi.	loop region
G275	ok	off	17	inaccurate	good	glycine, Ca^{2+} backbone ligand
N277	off	off	18	inaccurate	ambi.	unknown
G289	off	ok	17	inaccurate	good	glycine

The crystal structure indicates a weak hydrogen bonding of the backbone atoms with G237 that could potentially be different in solution, and the uneven number of residues in this region could explain a deviation from the database. Of the 44 torsion angles with the attribute “ambiguous”, only 25 showed a deviation of more than 30° for one or both torsion angles. These data

therefore show that most predictions with the attribute “good” are correct if they are not glycines. In fact, of 19 glycines in this construct, predictions were only accurate for 7. On the other hand, many predictions assigned as ambiguous showed actually an accurate prediction.

7.2.5 Predictions for individual domains in hN-1_{8-11}

TALOS+ was used to predict torsion angles for the hN-1_{8-11} construct. 155 predictions were made ranging from residue D295 to E450. 41 of these were assigned with the attribute “ambiguous” and G419 was predicted as dynamic. Figure 7.11 shows the predicted angles for individual domains EGF8-11 of the hN-1_{8-11} construct as compared to EGF12 for which torsion angles were derived from the crystal structure.

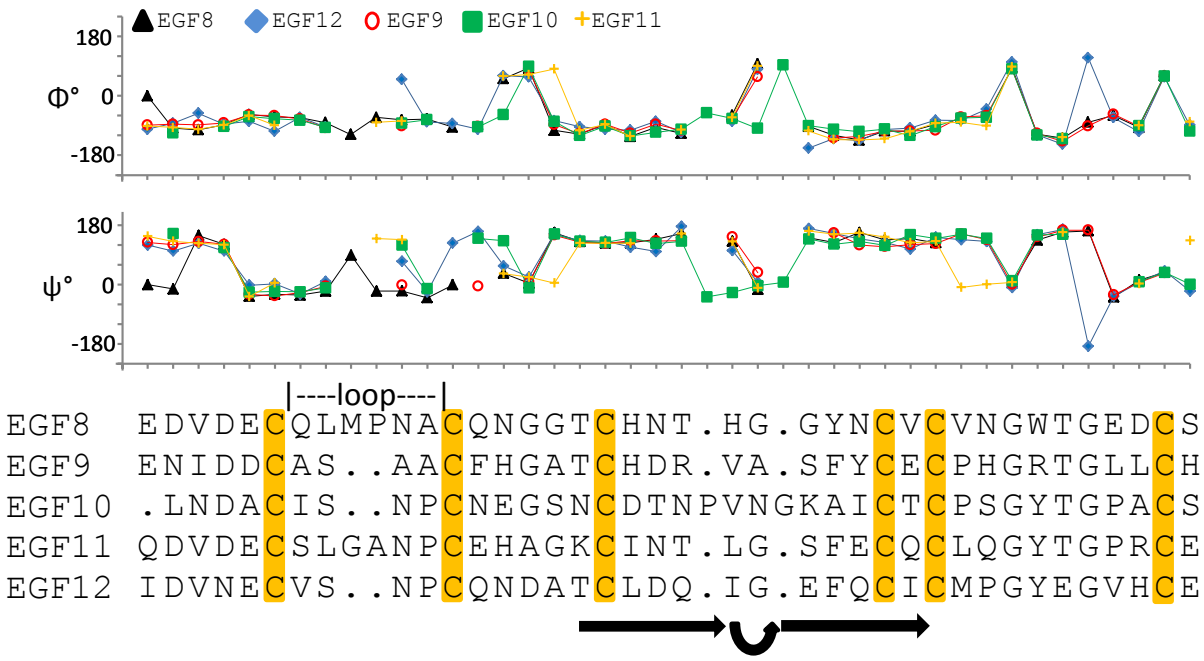


Figure 7.11 Prediction of hN-1_{8-11} torsion angles as compared to angles derived from the crystal structure of EGF12. Predictions are similar for residues around the conserved cysteines. Residues are aligned along the conserved cysteines (yellow). The arrows indicate the beta-sheet region.

23 of the ambiguous residues were taken out since the 10 matches of the database query were spread in different regions in the Ramachandran plot for these residues (not shown). These residues were Q307, V324, N325, S333, A343, C344, H346, S356, H371, N373, C381, G406, P407,

S417, L418, P422, C423, E424, N431, Y444, G446, P447 and C449. G419 and its neighbour A420 were taken out as well because of dynamics. Overall, the predictions matched the torsion angles derived from the crystal structure around the conserved disulphide-bonded cysteine residues suggesting that these are correctly formed. As in hN-$1_{11\text{-}13}$ and hN-$1_{4\text{-}7}$, the prediction for the glycine 6 residues after C5 was unsuccessful. However, good agreement can be seen for the conserved glycine 3 residues after C5. Interestingly, all predictions except for EGF10 matched the angles for the turn in the beta sheet. EGF10 is different here, as it has 10 residues in between C3 and C4 as compared to 8 residues for the other domains. The region after C2 did not show much overlap of the different predictions, although angles here were shown to be highly consistent in hN-$1_{4\text{-}7}$ and hN-$1_{11\text{-}13}$. The beginning of EGF8 shows deviations from the predictions of the other domains, which is consistent with its N-terminal position, a result in line with the data for the hN-$1_{11\text{-}13}$ construct. Since no X-ray structural information was available for EGF8, TALOS+ predicted torsion angles were compared with torsion angles from EGF11 in the hN-$1_{11\text{-}13}$ crystal structure (Figure 7.12).

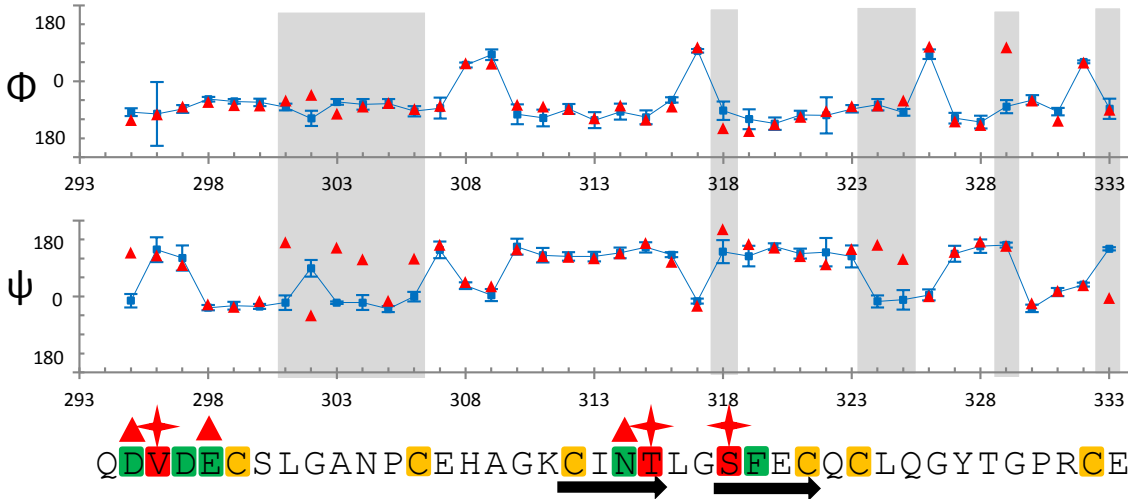


Figure 7.12 Overlay of prediction for EGF8 (blue circles) on the crystal structure of EGF11 (red triangles). The agreement is similar to the actual predictions of EGF11 (Figure 7.3). This suggests that EGF8 has a very similar structure compared to EGF11. Note that it is not known whether S318 is a backbone carbonyl ligand or not. Grey regions indicate where predictions are different from the torsion angles provided by the crystal structure of EGF11. Cysteines are highlighted in yellow, Ca^{2+} binding consensus residues in green, carbonyl backbone Ca^{2+} ligands in red; the red triangle indicates Ca^{2+} side chain ligands, and the red asterisk Ca^{2+} backbone ligands.

Both domains have the same number of residues in between cysteines and are both N-terminal and Ca^{2+} binding (Figure 7.12). Many of the predictions of torsion angles match quite well the torsion angles of EGF11, suggesting that EGF11 and EGF8 are very similar in structure.

EGF12 was used as a model to test the predictions of EGF9 (Figure 7.13). Similar to EGF8, the agreement was good, suggesting that the EGF8-EGF9 pair can be modelled by the EGF11-EGF12 crystal structure. Especially the predictions at the C-terminus are very precise.

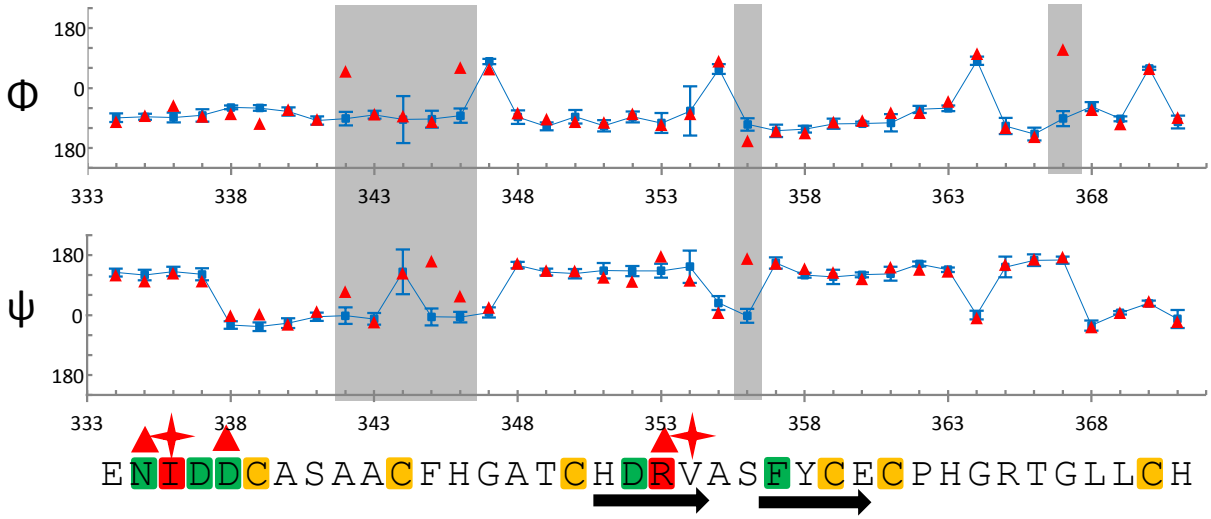


Figure 7.13 Overlay of prediction for EGF9 (blue circles) on the crystal structure of EGF12 (red triangles). The agreement is very good and suggests a high degree of conservation between EGF9 and EGF12. Grey regions indicate where predictions are not good. Ca^{2+} binding consensus residues in green, carbonyl backbone Ca^{2+} ligands in red; the red triangle indicates Ca^{2+} side chain ligands, and the red asterisk Ca^{2+} backbone ligands.

TALOS+ predictions for EGF10 were compared to torsion angles from available crystal coordinates of non-calcium binding domains (EGF4, EGF6, and EGF22) (Figure 7.14). None of these domains is expected to agree well with the torsion angles of EGF10, since EGF4 and EGF6 have a different number of residues between C3 and C4 than EGF10 and this region is distorted in EGF22. Similarly, these domains have some distinct types of amino acids at specific positions, such as a *cis*-proline in the loop of EGF6. Figure 7.14 shows an overlay over the different domains used. Whereas the number of residues between C1 and C6 were identical for EGF10

and EGF22, two blank residues were added after E159 in EGF4 and one blank residue was added after D238 in EGF6, which are located at the turn of the beta-sheet (arrow). The agreement was acceptable for many residues in all EGF domains.

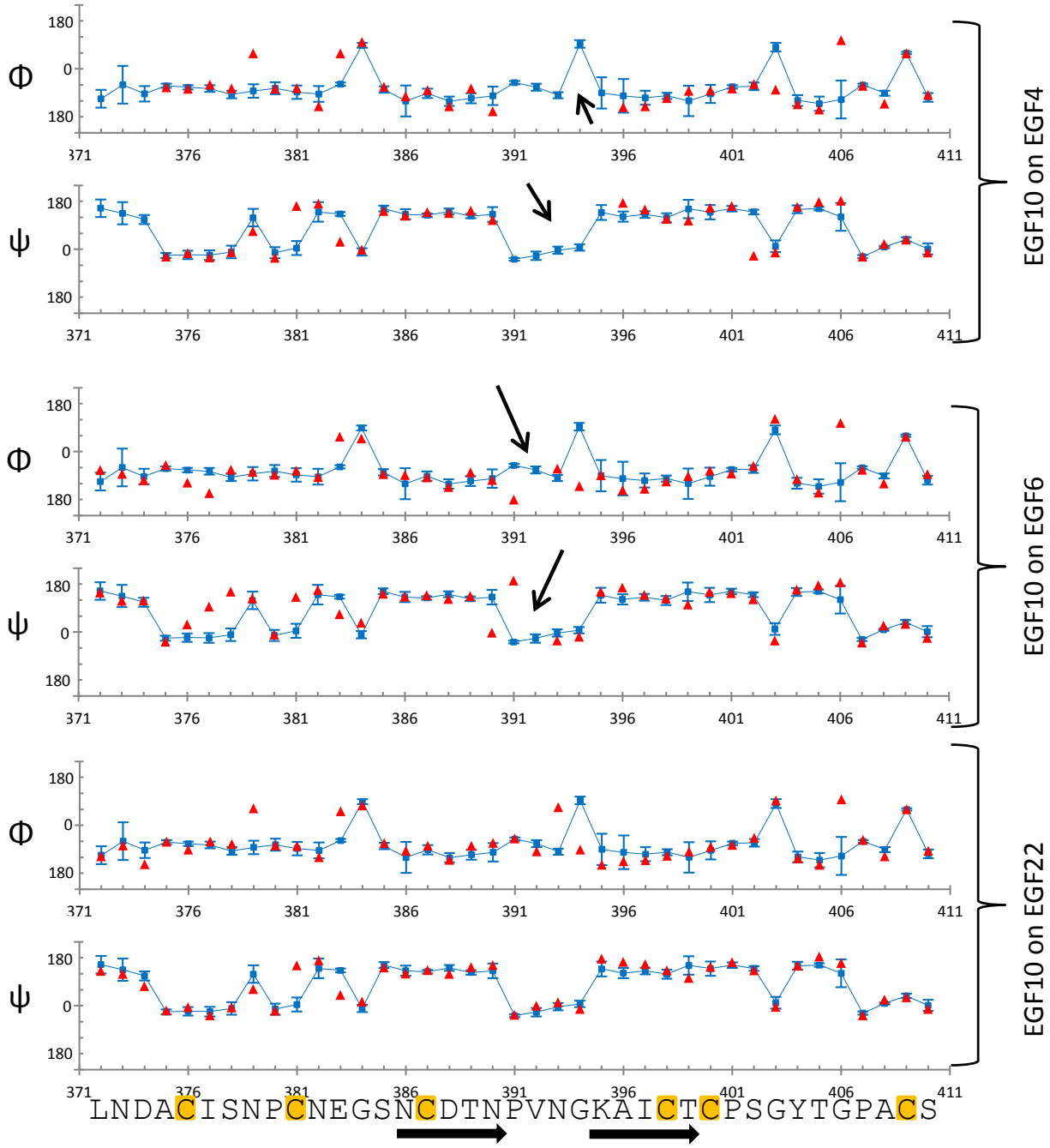


Figure 7.14 Prediction of torsion angles in EGF10 (blue circles) as compared to different available crystal structures (red triangles). In EGF4, one residue was added in the turn of the beta-sheet, however the electron density in the crystal structure was not good enough to define this region (arrows). For the modelling of EGF6, one blank residue was inserted after D328 to account for the extra residue in the turn of the major beta-sheet in EGF10 (arrows). Conserved cysteines are highlighted in yellow. The region of the major beta-sheet is indicated by black arrows.

As in the previous analysis, the prediction of glycine, three residues before C6 was not good. The predictions were in general better for the C-terminus of EGF10, suggesting that the formation of an interface to EGF11 is similar to other domains. Many psi angle predictions are not good at the beginning of EGF6. EGF22 is therefore probably the best model for EGF10. Further analysis, however, is required to test the interfaces formed by this domain; this will be shown in chapter 8.

TALOS+ predictions of EGF11 in hN-$1_{8\text{-}11}$ were compared to the torsion angles derived from EGF11 in the crystal structure hN-$1_{11\text{-}13}$ (Figure 7.15).

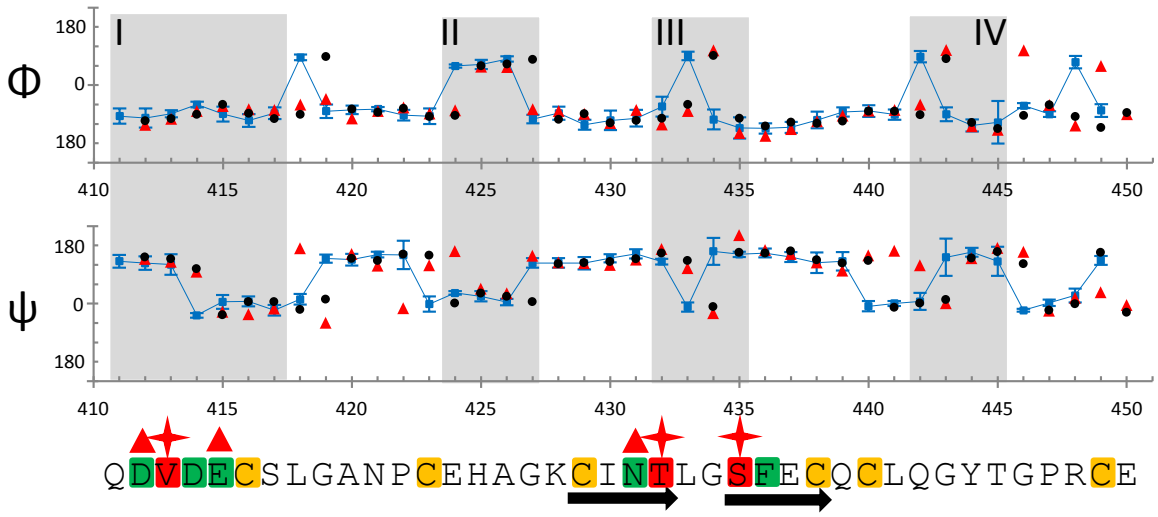


Figure 7.15 Phi/psi torsion angle predictions for EGF11. The predictions of EGF11 in the hN-$1_{8\text{-}11}$ construct (blue circles with error bars) vary between the ones predicted from the hN-$1_{11\text{-}13}$ dataset (black cicles). The torsion angles from the hN-$1_{11\text{-}13}$ X-ray structure are shown as a red triangle. Overall, the quality of the predictions is similar for both constructs. Ca^{2+} binding consensus residues in green, cysteines in yellow, carbonyl backbone Ca^{2+} ligands in red; the red triangle indicates Ca^{2+} side chain ligands, and the red asterisk Ca^{2+} backbone ligands.

Box I highlights residues that are at the N-terminus, these are slightly different and it is not clear whether the prediction are not good or whether the positioning of EGF10 in front of EGF11 has changed the torsion angles of this region. In box II, the phi angle prediction is worse for the hN-$1_{8\text{-}11}$ dataset, but the psi angle prediction is better. Box III shows residues that belong to the

turn of the major beta-sheet, these are probably slightly different because of the packing with EGF10. Interestingly, the angles of L433 and G434 showed a large deviation in opposite directions. Predictions are also very different for residues Q442 and G443, the residues before the consensus packing aromatic residue.

7.3 Secondary structure analysis

The addition of the neural network search in the TALOS+ version assigns each residue with a probability to adopt a secondary structure based on the ANN classification system. This is demonstrated on the structure of hN-$1_{11\text{-}13}$ (Figure 7.16). The height of the bars indicate the probability of a residue in question to be either sheet (blue) or helix (red), other residues are considered to belong to loop regions. All three major beta-sheets, and the first strand of the minor beta-sheet in EGF11 and EGF12 were identified. The second strand of the minor beta-sheet was not identified in EGF11 and EGF12, and both strands were not identified in EGF13.

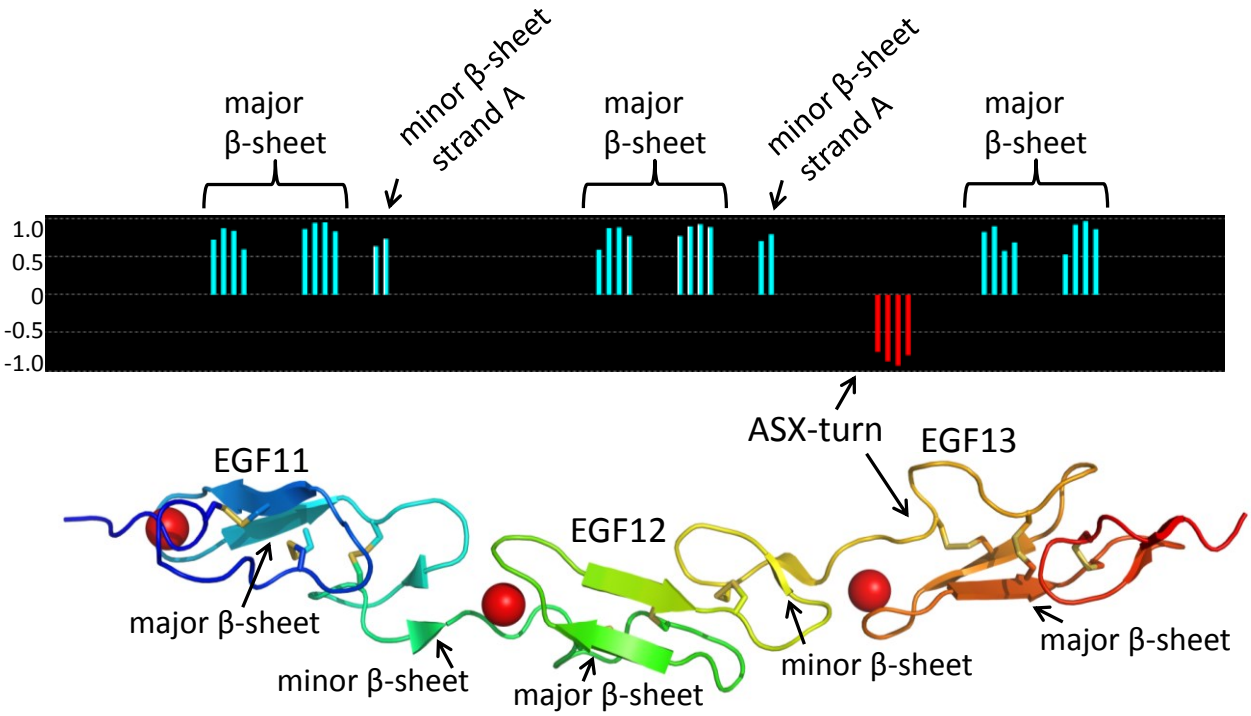


Figure 7.16 Secondary structure in hN-$1_{11\text{-}13}$. The prediction assigns each residue with a value from -1 to 1, for 1.0 the likelihood for the residue in question to be situated in a beta-sheet (blue) and for -1 helix (red). The last minor sheet was not predicted. The secondary structures were correctly predicted for EGF11 and EGF12 with two major beta-strands. However, the second strand of the minor beta-sheet was not identified. The ASX-turn of EGF13 was identified as the start of a helix.

A helix region was identified for the residue before C1 in EGF13 and three subsequent residues. This site contains the ASX motif (section 1.4.2) which initiates a helix-like turn, such as in EGF5 (section 6.5.1). Similar to hN-1_{11-13}, in hN-1_{4-7} major beta-strands were correctly predicted and the minor strands were not predicted very well (Figure 7.17).

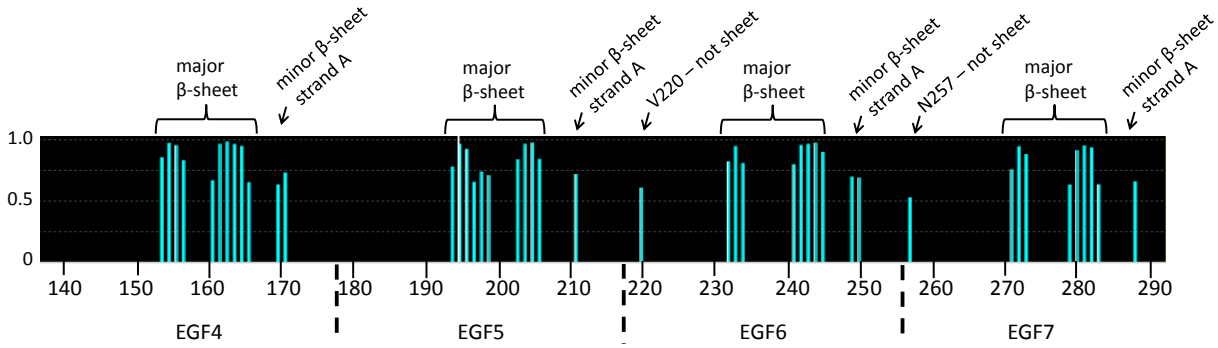


Figure 7.17 Secondary structure prediction of hN-1_{4-7}. Similar to hN-1_{11-13}, the predictions works well for major-beta-sheets, but not all residues involved in the minor-sheets are identified.

The minor-beta sheet at the end of EGF5, for example, is quite extensive and enables packing with the first residues of EGF6 (Figure 7.18). However, many assignments were missing in this region. The short helix at the beginning of EGF5 was not identified as here many residues weren't assigned.

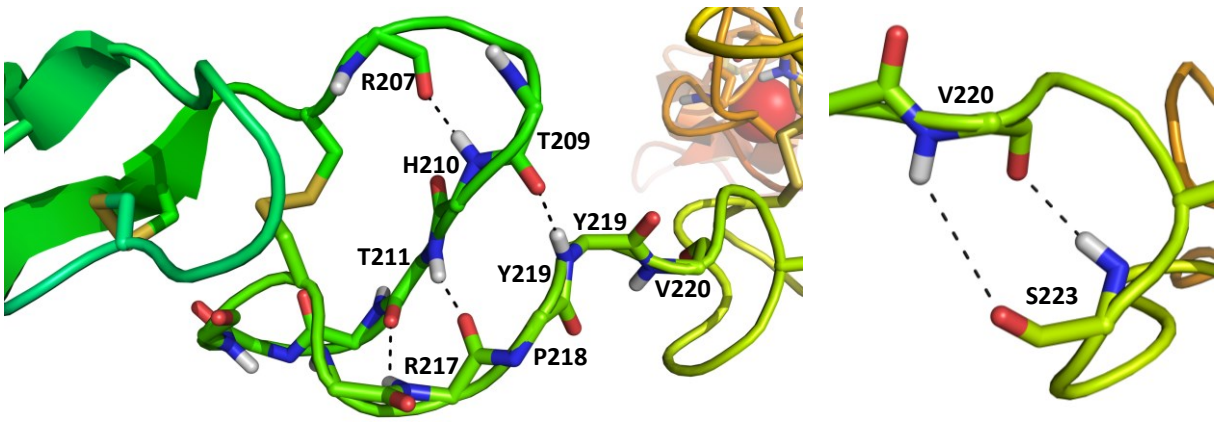


Figure 7.18 Minor beta-sheet at the EGF5-EGF6 interface. In the region after the major beta-sheet (EGF5), V220 was the only residue predicted to form a beta-sheet. It forms a beta-sheet with S223 (picture on the right). Probably due to the lack of assignments, S223 was not predicted to participate in this beta sheet. Similarly, the minor beta sheet (on the left) as exemplified by various residues was not predicted.

Since no crystal structure is available for hN-$1_{8\text{-}11}$, TALOS+ is used to predict secondary structure in this region (Figure 7.19). Predictions were made for all major beta-sheets, suggesting the formation of those in this construct. Whereas only five residues were predicted to form the major beta-sheet in EGF10, an analysis of NOE signals in a 3D NOESY dataset suggested connections between the backbone NH of N386 and T399, D388 and I397 and N390 and K395 (not shown). This, together with the observation of downfield shifted Hα shifts, suggests the correct formation of this beta-sheet. Helix residues were predicted for the region around C1 in EGF8 (residues E298-L301), EGF9 (C339 and A340) and EGF10 (I377), similar to the helix described in EGF5 (section 6.5.1). This region is known to form an ASX motif.

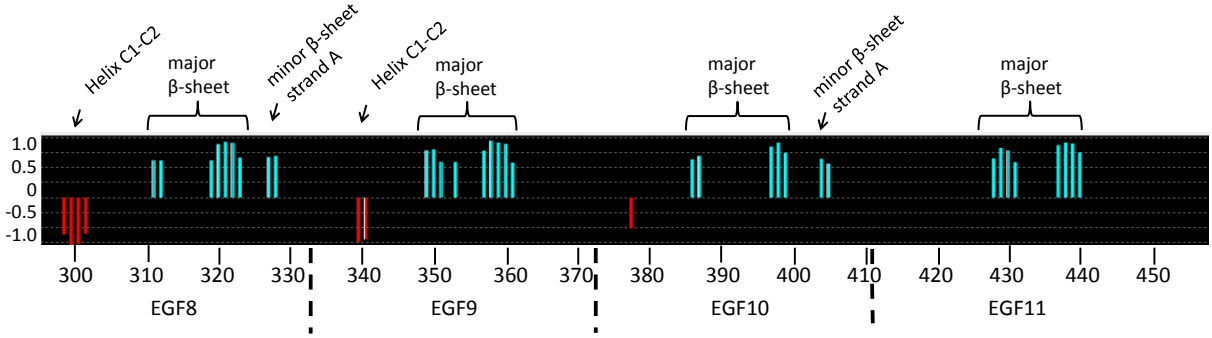


Figure 7.19 Prediction of secondary structure in hN-$1_{8\text{-}11}$. All major beta-sheets were predicted, as well as helical regions around C1 in EGF8 and EGF9.

7.4 Fast-timescale dynamics

The order parameter S^2 indicates the angular motion of an inter-nuclear vector, such as the NH backbone amide; it is 0 for an unhindered, completely free motion and 1 for a fixed orientation. Here, we have measured hetNOE values, which are sensitive to motions faster than the overall correlation time (<~10 ns). Furthermore, order parameters can be predicted from measuring the difference between a chemical shift and its random coil chemical shift, the so called secondary chemical shift. Random coil chemical shifts are defined as those shifts, measured at a very fast conformational exchange between all, energy weighted and theoretically possible conformations, in the unfolded state and without long-range contacts.

Hence, the proximity of a measured chemical shift to its random coil chemical shift, correlates with increased flexibility, (Berjanskii and Wishart, 2008). This has been used to estimate the state of protein disorder (Chou et al., 2002, Fiaux et al., 2002) and local flexibility. For example, a correlation between crystallographic B-factors and H_α shifts has been reported (Wishart and Sykes, 1994). However, since certain combinations of phi/psi torsion angles would, even in the absence of dynamics, be close to the random coil index, the correlation between dynamics and proximity to the random coil shift was limited (Berjanskii et al., 2002). The correlation was improved by adding the chemical shift of multiple nuclei – in addition to the H_α shift, the C_α, C_β, CO, N and NH shift were combined into a single parameter (Berjanskii and Wishart, 2005). Correlation was the best for the inverse of the averaged, weighted secondary shift, which was defined as the random coil index (RCI) and used in the TALOS+ order parameter predictions. The actual calculation of the RCI involves additional calculations steps, such as the smoothening of secondary shifts over adjacent residues and the chemical shift re-referencing; these steps are well documented (Berjanskii and Wishart, 2006).

In Figure 7.20, heteronuclear NOE measurements were compared to the predictions of S^2. Black arrows indicate order parameters below 0.7. TALOS+ was able to predict well residues with high mobility at the N- and C-terminus. In the publication of the hN-$1_{11\text{-}13}$ solution structure, hetNOE experiments indicated mobility at the loop between C1 and C2 in EGF11 (Hambleton et al., 2004). This was confirmed in the measurements described here, by both hetNOE measurements and S2 order parameter predictions, as indicated with an arrow and the letter “L” in the Figure. Similarly, the corresponding loop in EGF5 showed decreased S^2 values, but the hetNOE values were not as distinctive. This is the region where an α-helix was described in the crystal structure (section 6.5.1). Both EGF11 and EGF5 have five residues between C1 and C2. EGF8 also has five residues in this region, but the mobility of this loop could be confined because of the position of a proline four residues after C1. The other EGF domains had only 4 residues in this region.

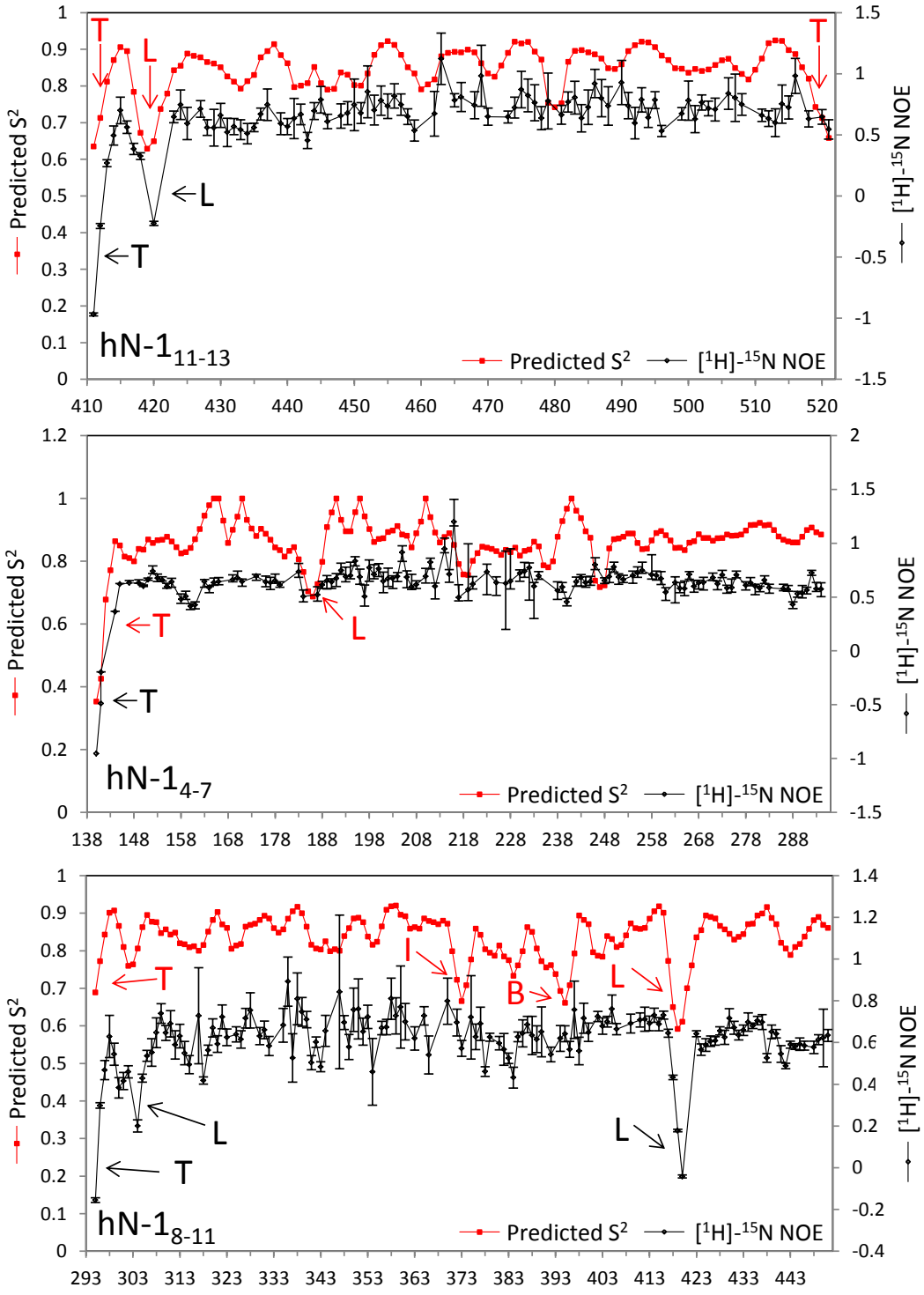


Figure 7.20 Predicted order parameter (S^2 in red) and measured [^{1}H]-^{15}N NOE values (black) for hN-1_{11-13} (top), hN-1_{4-7} (middle) and hN-1_{8-11} (bottom). Red arrows indicate regions where the predicted order parameter was less than 0.7. This was the loop (L) between C1 and C2 in EGF5 (hn-1_{4-7}) and EGF11 in both the hN-1_{11-13} and hN-1_{8-11} construct, as well as several residues at the termini (T). In hN-1_{8-11}, residue N373 at the interface (I) between EGF9 and EGF10 had a predicted order parameter of 0.666, suggesting loose inter-domain contacts. Furthermore, residues G394 and K395 at the turn of the major beta-sheet (B) in EGF10 had lower predicted S^2 values as well. Similarly, regions with a very low heteronuclear NOE value are indicated with black arrows and correspond to loop regions and residues at the N-terminus only. The very flexible residues of the birA-tag at the C-terminus of hN-1_{11-13} were removed.

In hN-$1_{8\text{-}11}$, N373, a residue between C6 of EGF9 and C1 in EGF10 and residues at the turn of the major β-sheet in EGF10 (G394, K395) had lower predicted S2 values as well. However, this was not evident from the hetNOE values, although these showed a large error within these regions. The reduction in order parameter at the linker region between EGF9 and EGF10 is interesting, since this region was suggested to provide flexibility because of the lack of Ca^{2+} binding (Cordle et al., 2008b). This finding is therefore in line with this suggestion, the fact that there is no conserved aromatic packing residue, and the lack of success during crystallography. The fact that the hetNOE values weren't decreased in this region could suggest flexibility on a slower timescale; this will be addressed in chapter 8 in a residual dipolar coupling analysis (RDC).

7.5 Ca^{2+} backbone ligands

In section 6.7 it was argued that EGF-cbEGF domain pairs are around ~60° more twisted than cbEGF-cbEGF pairs if the residue at three positions before C4 provides a backbone carbonyl Ca^{2+} ligand. A method to detect the Ca^{2+} binding property of this residue could therefore give long-range structural information. Chemical shift differences were previously used to monitor Ca^{2+} ligands in EF-hand proteins (Biekofsky et al., 1998) where Ca^{2+} binding to the carbonyl of residue 7 caused a chemical shift change of the amido nitrogen (^{15}N nucleus) in residue 8 of +4 ppm to +8 ppm (Figure 7.21). This was interpreted as a de-shielding effect of the amido group as a result of the stronger polarisation in the backbone carbonyl group of amino acid 7. Here, the effect of Ca^{2+} binding to the backbone carbonyl of residue (i) on nuclei C'(i), N(i+1) and HN(i+1) is measured (Figure 7.21). Table 7.3 shows the values for chemical shifts as adjusted by TALOS+; this adjustment removes the random coil chemical shift of the specific residue type (Berjanskii and Wishart, 2008) and allows comparison of structural properties independent of the exact amino acid sequence. Backbone carbonyls of the residue three positions before C4 had adjusted chemical shifts of -1.40 ppm in EGF5 of hN-$1_{4\text{-}7}$ and -1.48 ppm in EGF11 of hN-$1_{11\text{-}13}$ which were both described as Ca^{2+} ligands.

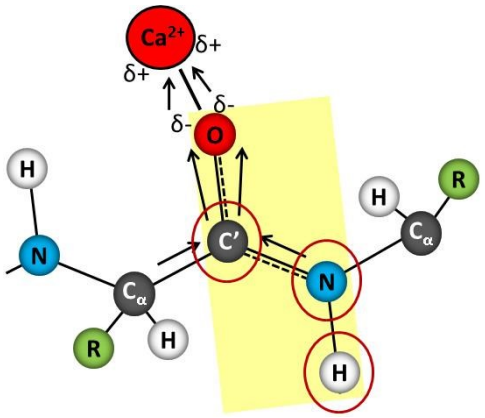


Figure 7.21 Peptide bond and residues possibly affected by Ca^{2+} binding. The carbonyl bound to Ca^{2+} is expected to show an increased electronegativity. This can cause a de-shielding of covalently attached atoms, such as the C', N and HN (red circles).

These were significantly different to -3.74 ppm, -4.58 ppm and -4.84 ppm, the values obtained for EGF7, EGF12 and EGF13 for which this residue is not a Ca^{2+} ligand.

Table 7.3 Adjusted chemical shifts of residues tested for Ca^{2+} binding.

	Ca^{2+} backbone ligand (ppm)		Not a ligand (ppm)			Unknown (ppm)		
domain	5	11 (11-13)	7	12	13	8	9	11 (8-11)
C'(i)	-1.40	-1.48	-3.74	-4.58	-4.84	-2.59	-3.83	-1.32
N(i+i)	-4.55	-4.17	-8.70	-8.92	-7.54	-0.30	-10.35	-5.86
HN(i+1)	0.57	0.58	-0.51	-1.30	-0.72	-0.24	-0.66	0.65

The adjusted chemical shift of the backbone nitrogen in the residue i+1 had a shift of around -4 ppm or -5 ppm for residues in EGF5 and EGF11 (hN-1_{11-13}), which was significantly more negative for EGF7, EGF12 and EGF13 with values around -8 ppm or -9 ppm. Similarly the HN of the residue i+1 had shifts of 0.5-0.7 ppm for EGF5 and EGF11 (hN-1_{11-13}) which were quite different to values between -0.5 ppm and -1.3 ppm for EGF7, EGF12 and EGF13. All measurements are summarised in Figure 7.22A, clearly showing two categories of chemical shifts. Adjusted chemical shifts for EGF11 in hN-1_{8-11} (Figure 7.22B, red) matched quite well the chemical shifts obtained for EGF5 in hN-1_{4-7} and EGF11 in hN-1_{11-13} and confirmed the residue in question as a backbone carbonyl ligand in the hN-1_{8-11} construct.

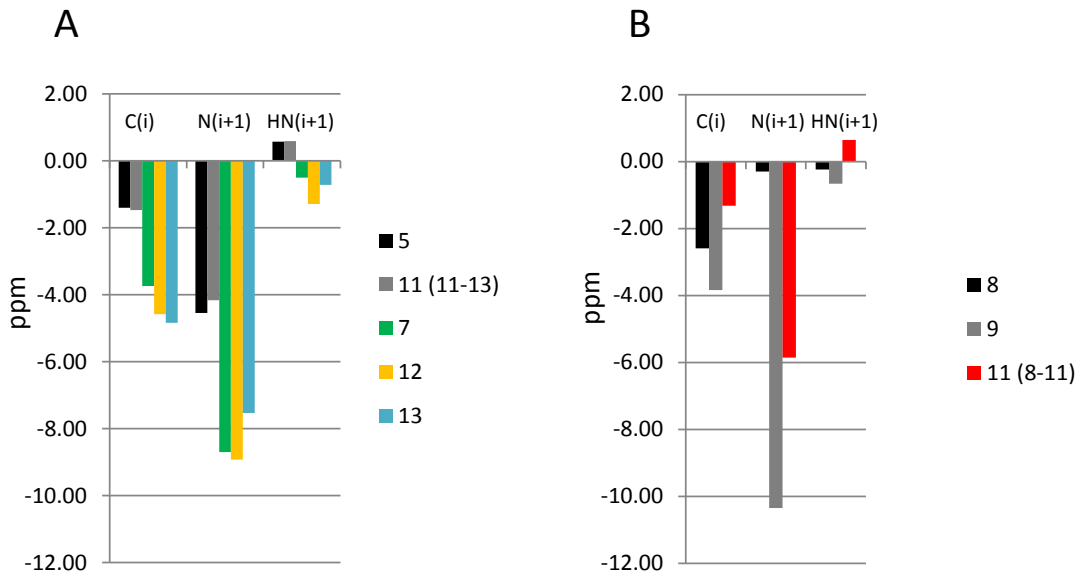


Figure 7.22 Adjusted chemical shifts of EGF5/EGF7 and EGF11 of the hN-1_{11-13} construct (A) and EGF8, EGF9 and EGF11 of the hN-1_{8-11} construct (B). In (A) adjusted C(i), N(i+1) and HN(i+1) values for EGF5 (black) and EGF11 (grey) agree well with each other, both of which represent Ca^{2+} binding in the residue at position three before C4. Non-Ca^{2+} ligands show different values as seen for EGF7 (green), EGF12 (yellow) and EGF13 (blue). In (B), adjusted chemical shifts of constructs without structural information are displayed. Whereas EGF11 of hN-1_{8-11} (red) agrees with Ca^{2+} binding, EGF9 agrees with EGF domains in which this residue is not a ligand. For EGF8, no conclusion can be drawn, since the values were in between both categories, the reason for this is unclear.

Values for EGF9 on the other hand correspond well to the category in which the residue in question is not a ligand. Values for EGF8 did not match either of the categories. The chemical shifts for C'(i) and HN(i+1) of EGF8 are intermediate between values observed for the two categories but the correct 15N(i+1) shift is close to 0 ppm. This latter shift must be influenced by other factors in EGF8. These results, together with the observations of the previous chapter, suggest interfaces similar to EGF4-5 and EGF22-23 for EGF10-11 and a similar interface for EGF8-9 when compared to EGF11-12 or EGF12-13.

7.6 Discussion

TALOS+ predictions are shown to be highly accurate for many residues in EGF-like domain constructs, a finding that is based on results of the hN-1_{11-13} and the hN-1_{4-7} crystal structure. Exceptions are the flexible loop and the residue type glycine.

The predictions based on TALOS+ could be improved by the addition of the chemical shifts measured during this thesis to the TALOS+ database. Often Ca^{2+} backbone ligands and the specific glycine, three residues before C6, are consistently wrongly predicted. Since the overall EGF-like fold is highly conserved with very similar torsion angles, the question arises to whether TALOS+ can differentiate between the characteristics of individual EGF domains. This has been shown successfully for EGF6 that has a significantly different set of torsion angles at the C-terminus, when compared to EGF4, EGF5 and EGF7. Moreover, many residues at the interface, such as the residues at the beginning of EGF7 were predicted well. However, the error for these predictions is not very much lower than the difference when comparing different torsion angles.

TALOS+ torsion angle predictions showed that EGF8-EGF9 is structurally very similar to EGF11-EGF12 and this information is helpful for further analysis (chapter 8). Similarly, the interface between EGF10-EGF11 seems to be comparable to other interfaces.

TALOS+ was able to detect beta-sheet secondary structures for all major beta-sheets. Many minor beta-sheets were not predicted well, often because of the lack of assignments. Furthermore, using adjusted chemical shifts from TALOS+, it has been possible to differentiate Ca^{2+} binding to the backbone carbonyl at position three before C4 which was linked to an increased twist angle.

Therefore, TALOS+ is able to give detailed structural information for systems with complete resonance assignments. Interestingly, one of the interface residues between EGF9-EGF10 showed a low order parameter prediction of <0.7. This is in line with the prediction that EGF9-EGF10 is not rigid because of the lack of the consensus aromatic packing residue in EGF9. The interface formed by EGF9-EGF10 will be analysed in chapter 8.

8 Defining inter-domain orientations using residual dipolar couplings

8.1 Introduction

Little is known about the structural arrangement of highly conserved calcium-binding and non-calcium binding EGF domains of Notch on the cell surface to enable ligand binding and receptor activation. Whereas previous deletion experiments in cell aggregation assays with overexpressed Notch and ligand indicated that EGF11-12 are necessary and sufficient for ligand binding, genetic experiments clarified that this minimal region is not sufficient for Notch signalling (Rebay et al., 1991, Rebay et al., 1993). Similarly, cell assays using overexpressed ligands and incubation with isolated Notch fragments confirmed the major contribution of EGF11-12 to ligand binding, but a significantly reduced binding (<50%) was found when EGF domains 6-10 or 24-36 were deleted (Lei et al., 2003, Xu et al., 2005). Because EGF6, EGF10, and EGF22 were predicted to be non-calcium binding EGF domains, some flexibility at these sites was proposed resulting in a non-linear organisation at the cell surface (Hambleton et al., 2004) (Figure 8.1).

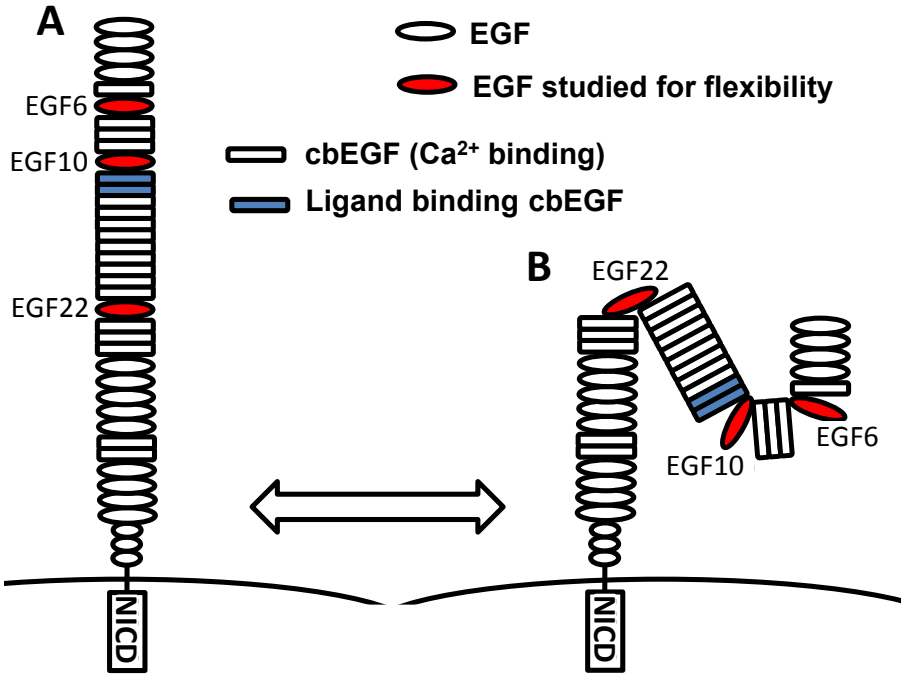


Figure 8.1 Proposed structural models for the Notch receptor. The cell-surface organisation of Notch-1 is unknown. Non-calcium binding EGF6, EGF10 and EGF22 were predicted to confer flexibility to the receptor and arrange Notch-1 in a jack-knife (B) organisation as opposed to a linear arrangement (A).

In Chapter 4, the rigidifying effect of Ca^{2+} on tandem cbEGF domains was studied by measuring Ca^{2+} dissociation constants for the N-terminal region of Notch. In Chapter 6, novel crystal structures for hN-1_{4-7} and hN-1_{21-23} displayed 2 examples of rigid interfaces formed by non-calcium binding EGF domains. In contrast to the jack-knife model (Figure 8.1B), EGF21-22 form a rigid and elongated structure, although a bend, consistent with the model was observed between EGF5-6. Whereas the information obtained by crystallography for EGF4-7, EGF11-13 and EGF21-23 indicates rigidity for the Notch receptor, it is unclear if there are other, more flexible, sites in the Notch receptor and whether constructs including these sites have less potential to crystallise. One of these potentially flexible regions could be around non-calcium binding domain EGF10; interestingly crystallisation attempts for constructs hN-1_{9-11}, hN-1_{8-11}, hN-1_{9-10}, hN-1_{9-13} and hN-1_{10-13} all proved unsuccessful. Similarly, attempts to refold the large non-calcium binding region at the N-terminus, hN-1_{1-4} and hN-1_{1-5} were unsuccessful. In chapter 7, chemical shifts were used to predict backbone torsion angles in solution that were consistent with the domain structures observed by X-ray crystallography for hN-1_{4-7} and hN-1_{11-13}. In addition, the prediction of the random coil index by TALOS+ indicated lower S^2 order parameters for the residues in between C6 of EGF9 and C1 in EGF10, suggesting flexibility at the interface region. Although a loose interface between EGF9 and EGF10 was not obvious from hetNOE measurements, interdomain motions could be a result of slower-timescale (μs-ms) motions that are not measurable by hetNOE. The characterisation of motions within the Notch receptor is important since large-scale domain motions are often the basis of many biological processes such as enzyme catalysis and the recognition of binding partners (Gerstein et al., 1994, Anderson et al., 1979, Boehr et al., 2006). Therefore, the events leading to the formation of the Notch-ligand complex, as well as the subsequent cleavage of the Notch receptor and its transport into different cell compartments could be a result of large domain rearrangements.

In this chapter residual dipolar couplings (RDC) are used to describe the domain orientation and rigidity of the regions hN-$1_{4\text{-}13}$. Whereas NOE's or J-couplings provide short-range information, RDCs provide long-range information, for a review see Lipsitz and Tjandra (2004). This is particularly helpful for the analysis of EGF domains as these show little secondary structure or interdomain contacts.

8.2 RDC measurements

8.2.1 Theory

Dipolar couplings arise from the interaction of a pair of spins with the electromagnetic field, a physical property used in solid-state NMR. Equation (1) shows that the dipolar coupling (D) is dependent on the inter-nuclear distance, the term K, and on the angle θ between the internuclear vector and the magnetic field.

$$D = \frac{\kappa}{R^3}(\cos^2\theta - \frac{1}{3}) \qquad (1)$$

R is the distance between the coupled nuclei; in the case of the ^{1}H-^{15}N RDCs measured here, this distance is ~1.04 Å. The term K is a constant, defined as

$$\kappa = -\frac{2}{8\pi^2}\gamma_I\gamma_S\mu_0\hbar \qquad (2)$$

This term consists of the gyromagnetic ratios γ of the spins I and S, the permeability of vacuum μ_0, and Planck's constant ħ. Thus, the size of the coupling can be determined if the angle θ between the internuclear vector and the magnetic field is known. Since the direction of the magnetic field is constant, the important structural information resulting from dipolar couplings are the angles between the internuclear vectors within a single molecule. In an isotropic medium, such as in aqueous solution, the dipolar couplings average to zero due to molecular tumbling and the information provided by the dipolar coupling is lost. However, molecules can

be partially aligned in a liquid crystalline medium through steric or electrostatic forces and the observed coupling is averaged over time, hence the name residual dipolar coupling (3).

$$\overline{D} = \frac{\kappa}{R^3}(\overline{\cos^2\theta} - \frac{1}{3}) \qquad (3)$$

The solution for this equation can be visualised by the *probability tensor* P that describes the probability to find each of the three axis of the coordinate frame in a molecular structure (x, y and z) along the external magnetic field (Figure 8.2).

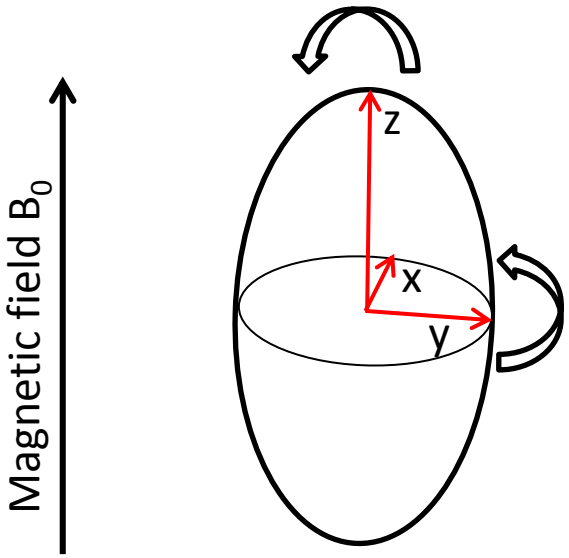


Figure 8.2 Probability tensor visualised by an ellipsoid. The shape of the ellipsoid indicates the probability of alignment in each of the three axes defined by the molecular frame (x, y and z). Here, the direction of alignment coincides with the direction of the magnetic field (B_0).

The dipolar coupling can be calculated if the Cartesian components of the vector of interest and their probability of alignment are known (4) (Kramer et al., 2004).

$$\overline{\cos^2\theta} = P_x {r_x}^2 + P_y {r_y}^2 + P_z {r_z}^2 \qquad (4)$$

However, the alignment tensor is usually described in its traceless form, in which the three principal components, A_{xx}, A_{yy} and A_{zz} sum to 0 instead of 1 (Lipsitz and Tjandra, 2004). This representation is more difficult to visualize as at least one of the components must be <0.

The observable residual dipolar coupling (D) of nuclei A and B in solution, is described (5, 6).

$$D_{NH} = \sum -S\left(\frac{\mu_0 h}{8\pi^3}\right)\gamma_N\gamma_H\langle r_{NH}^{-3}\rangle cos^2\Phi_i \; A_{ii} \qquad (5)$$

Where ϕ_i is the angle between the NH bond, and the principal axis of the molecular alignment tensor, and S is the order parameter. This can be recast in terms of an axial, D_a and a rhombic D_r component of the alignment tensor.

$$D_{NH}(\alpha,\beta) = D_a \; \langle(3\cos^2\alpha - 1) + \frac{3}{2}\mathrm{R}(sin^2\alpha cos2\beta)\rangle \qquad (6)$$

$$D_a \; = \left[A_{zz} - \frac{A_{xx}+A_{yy}}{2}\right]/3 \qquad (7)$$

$$D_r \; = \left[A_{xx} - A_{yy}\right]/3 \qquad (8)$$

$$R = \; D_r/D_a \qquad (9)$$

Here, α is the angle between the NH vector and the z-axis of the tensor and β describes the position of the projection of the NH vector in the x-y plane. Therefore, the calculation of the alignment tensor requires the determination of five variables, D_a, R, θ, ϕ and ψ; θ, ϕ and ψ describe the orientation of the alignment tensor in the molecular frame (Euler angles).

8.2.2 RDC measurements

The potential for the measurement of dipolar couplings was first described by Alfred Saupe (1963). The introduction of liquid crystalline media, such as bicelles (Sanders and Prestegard, 1990, Sanders and Schwonek, 1992) or mixtures of polyethylene glycol and hexanol (Ruckert and Otting, 2000), phage or DNA and the intrinsic paramagnetic property of certain proteins (Tolman et al., 1995), led to the new field of RDC measurements for the calculation of protein structures. In the study of the Notch constructs, the use of phage was not possible since the concentrations of Ca^{2+} required to saturate cbEGF domains led to the aggregation of phage

particles. DMPC/DHPC bicelles were tested on the hN-$1_{11\text{-}13}$ construct – even at the lowest concentration of bicelles necessary to form a liquid crystalline medium (~3%), the alignment was too strong and signals couldn't be measured because of severe line broadening. Alignment was successful using a combination of hexanol and alkyl-poly(ethylene glycol) (Ruckert and Otting, 2000) (described in section 2.10). In isotropic solution, the splitting in the ^{15}N dimension of an un-decoupled HSQC corresponds to the one-bond ^{1}H-^{15}N coupling $^{1}J_{NH}$. In alignment medium, this splitting corresponds to the sum of $^{1}J_{NH}$ and the ^{1}H-^{15}N RDC. The RDC values can be determined by subtracting the splittings measured in these two spectra. The doubling of peaks as a result of the splitting leads to spectral crowding. Therefore, two spectra are recorded in which the upfield and the downfield components are collected separately. Here, the IPAP (in-phase/anti-phase) method was used (Ottiger et al., 1998) (Figure 8.3).

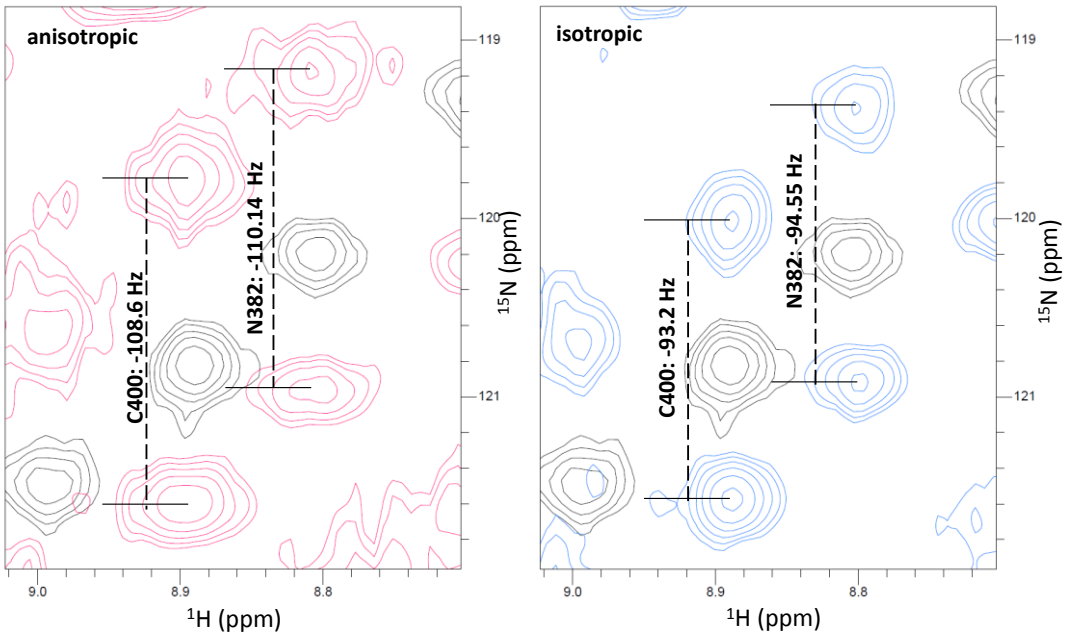


Figure 8.3 Measurement of RDC using the in-phase/anti-phase (IPAP) pulse sequence. The HSQC peak is indicated in black. Splitting during the isotropic measurement correspond to J-coupling (blue on the right). After adding alignment media (left), the splitting (red) is the sum of the J-coupling and the residual dipolar coupling.

The difference between both datasets corresponds to the RDC only, this is -15.4 Hz for C400 and -15.6 Hz for N382. In all measurements, the percentage of alignment medium was

around 2-2.5% peg/hexanol, and the D_a values around 9-20 Hz. Too much alignment medium would lead to line broadening and too little alignment wouldn't allow the formation of the lyotropic liquid crystalline phase. Similarly, the relative error involved in the determination of the correct peak position is more significant for small RDC values. The degree of alignment was measured by the D_2O residual quadrupolar splitting (Figure 8.4).

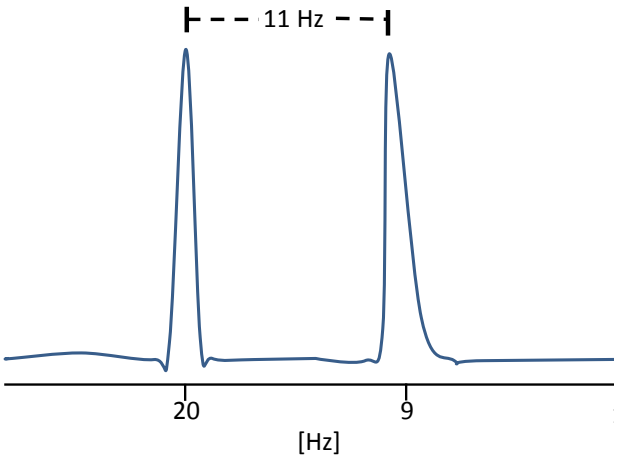


Figure 8.4 D_2O residual quadrupolar splitting. A splitting of 11 Hz is observed. The quadrupolar coupling is a result of the addition of alignment media and the splitting in Hertz correlates with the strength of alignment.

From the various parameters obtained from fitting RDCs, the overall alignment of the molecules can be used to infer information about the shape and charge of the molecule in question (Figure 8.5). Long and rod-like molecules will align stronger than molecules that are round and sphere-like.

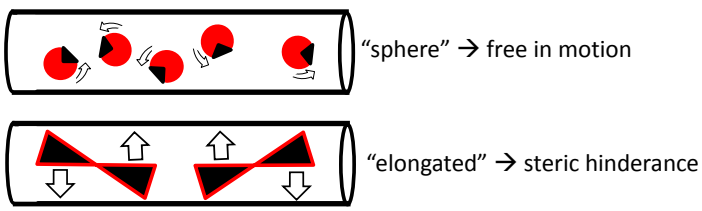


Figure 8.5 Alignment mechanisms. Proteins, when coming close to the surface of the alignment medium (black tube), cannot adopt all conformations. This depends on their shape – sphere-like molecules will not align in such a medium, whereas elongated molecules will have a "preferred orientation".

Recently, the computer software PALES has been developed to predict the alignment strength from the shape of the molecule (Zweckstetter, 2008). For a very flexible molecule, the alignment will be smaller than for a very rod-like molecule. Therefore, the measured alignment strength can be used to estimate degrees of flexibility. Preferred alignment media for the prediction of tensors are bicelles and peg-hexanol systems because of a mostly steric interaction between alignment medium and protein. Domains connected by a flexible linker are predicted to align less than the same domains rigidly attached to each other. In a recent study, two identical SH3 domains were connected by linkers of different length and correlated with predicted alignment parameters using PALES (Yuwen et al., 2011, Zweckstetter and Bax, 2000).

8.2.3 Domain orientation

In this study, domain orientations are determined as described here. ^{1}H-^{15}N RDCs are measured for the different EGF-domain constructs. For each individual domain, either the crystal structure or a homology model is used to fit an alignment tensor using a computer program written by Prof C Redfield. This program uses equation (5) and fits the five components of the alignment tensor, Da, R, theta, phi and psi, minimising the resulting error (10) which is defined as the sum of the squared differences between the experimental and calculated RDCs.

$$Err = \frac{\sum\left(D^{meas} - D^{calc}\right)^2}{n} \qquad (10)$$

Flexible residues, as shown by low heteronuclear NOE values (NOE ratio <0.45) were removed from the fit since their position may be distorted in the structure or the motion will scale down the RDC (equation 5). The alignment tensor is fitted in an iterative way – after an initial fit, a small number of residues with very poor agreement (expt-calc)^2 > 50 Hz may be removed as these would distort the correct determination of the alignment tensors. Large errors, measured for the excluded residues, can be a result of a distortion of the NH vector within the structure,

because of crystal contacts, or because of mobility with the lack of electron density as a result. The fitting procedure is then repeated until a good fit is obtained for each of the individual EGF domains. The errors in the fitted D_a, R and angles are estimated in two different ways. First, the fitting is repeated 500 times using a Monte Carlo approach in which random errors taken from a Gaussian distribution with a standard deviation of 2 Hz or 4 Hz are added to each RDC value and a set of D_a, R and angles fitted. The reported alignment parameters and errors represent the average values from 500 runs +/- the standard deviation of these 500 fits. A second error analysis was adopted to investigate the dependence of D_a, R and angles on the set of RDCs values used. Therefore, fits were repeated 100 times and 4 residues, accounting for around 15-25% of the residues per EGF domain, were removed randomly in each fit. In the analysis, the q-value is shown as an additional estimate of the quality of the fit, this is an established parameter that is commonly used and indicates the quality of the fit; a q-value of 0 indicates perfect agreement (Cornilescu et al., 1998).

$$q = \frac{\sqrt{\sum(D^{meas} - D^{calc})^2}}{\sqrt{\sum(D^{meas})^2}} \qquad (11)$$

This procedure is applied to all individual EGF domains, for which RDCs were collected. Assuming a rigid connection between the individual domains of a construct, all domains will tumble and align as a single unit. Therefore, the strength of alignment (D_a and R) and angles will be the same for all domains. The relative orientation of domains observed in crystal structures or homology models can be tested by fitting the RDCs for pairs of domains using single D_a, R, theta, phi and psi values. A poor fit indicates that the relative domain orientation in the model structure does not represent that observed in solution. In order to define the domain orientation in solution, the RDC values for a pair of domains is fitted using single D_a and R values, but two sets of theta/phi/psi angles which represent the orientation of each individual domain in the alignment medium. These angles can then be used to build a model of the domain pair. Because

of the cosine square of the angles in equation (5), 4 sets of angles based on the rotation around each axis, will fit a set of RDCs. However, the covalent connections between EGF domains, steric clashes and the availability of NOEs and models, makes it often possible to reduce the possible set of angles to a single orientation. If very different D_a and R values are obtained for EGF domains in a construct, this can be explained by the fact that the two domains are not rigidly connected; they don't tumble as a single unit and will experience different strengths of alignment. This is likely to be a result of mobility in the linker between domains.

The program Module can also be used to define domain orientations (Dosset et al., 2001). However, this approach allows D_a and R to differ for the two domains. If a rigid structure exists for two domains, then the RDCs should be described by a single set of D_a and R value. Therefore, in-house software written by Professor C Redfield was used for the fitting procedure.

An example for the determination of interdomain orientation was published in 2001 in which the position of the EF hand in the N-terminus of calmodulin in solution was compared to that observed in the crystal structure (Chou et al., 2001). With a large set of RDCs and J-couplings, the authors were able to measure precisely the angles between an EF hand and the remaining molecule that was significantly different to the crystal structure. Figure 8.6 illustrates this method. Since the alignment tensor is difficult to visualise as at least one of its components is negative, eigenvectors for the three different angles are presented. The domain orientation of two domains (depicted in green and grey) towards each other is unknown. Fitting individual alignment tensors show three eigenvectors, x, y and z; the length of these vectors correspond to the probability of alignment in each direction and contain therefore information for D_a R. The fitting of the alignment tensor shows that the length of the three axes x,y and z is identical for both domains, but the vectors point in different orientations. Both domains are then rotated, until the directions of the three vectors align.

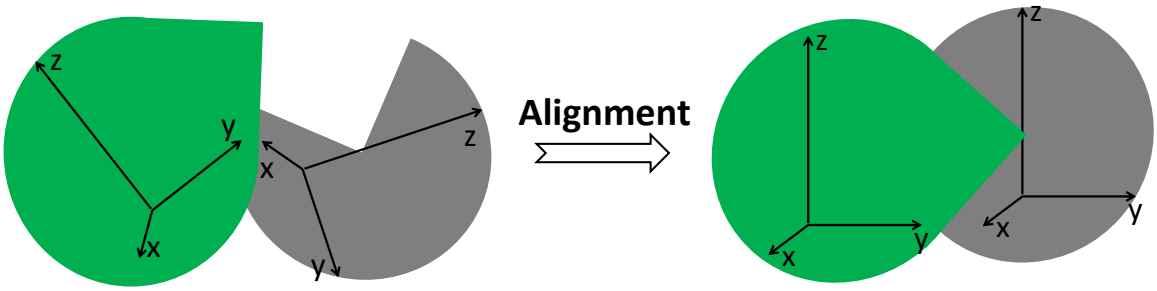


Figure 8.6 Alignment in a 2 dimensional space. (Left) alignment tensors are fitted to individual domains. (Right) the three eigenvectors are then aligned so that they point in identical directions.

The distance between the two domains, however, cannot be determined by RDC measurements and must therefore be measured with other methods such as NOE measurements. Previously, RDCs have been used to refine the solution structure of the EGF-AB pair within the LDL receptor (Saha et al., 2001). Here, the analysis of ^{15}N-^{1}H RDC measurements on different constructs encompassing the region hN-$1_{4\text{-}13}$ is presented. The aim is to produce a low-resolution structure of this region and to estimate the potential degree of domain flexibility in solution by comparing the strength of alignment of different domains. The use of RDCs to define domain orientations for EGF-containing constructs was first tested on the hN-$1_{11\text{-}13}$ construct, for which both a crystal and a solution structure are available (Hambleton et al., 2004, Cordle et al., 2008a). In addition, the D469G was shown to abolish Ca^{2+} binding in EGF12, leading to a less defined interface between EGF11 and EGF12. This mutant therefore presents a good test system containing a site of flexibility.

8.3 Analysis of the hN-$1_{11\text{-}13}$ domain architecture

8.3.1 RDC measurement in hN-$1_{11\text{-}13}$

RDC values for hN-$1_{11\text{-}13}$ were measured in ~2.0% peg/hexanol, 30 mM Ca^{2+} and 1mM protein concentration, resulting in a D_2O splitting of 7.1 Hz. The RDCs were large as expected for an elongated structure and vary between -30 Hz and +30 Hz (Figure 8.7).

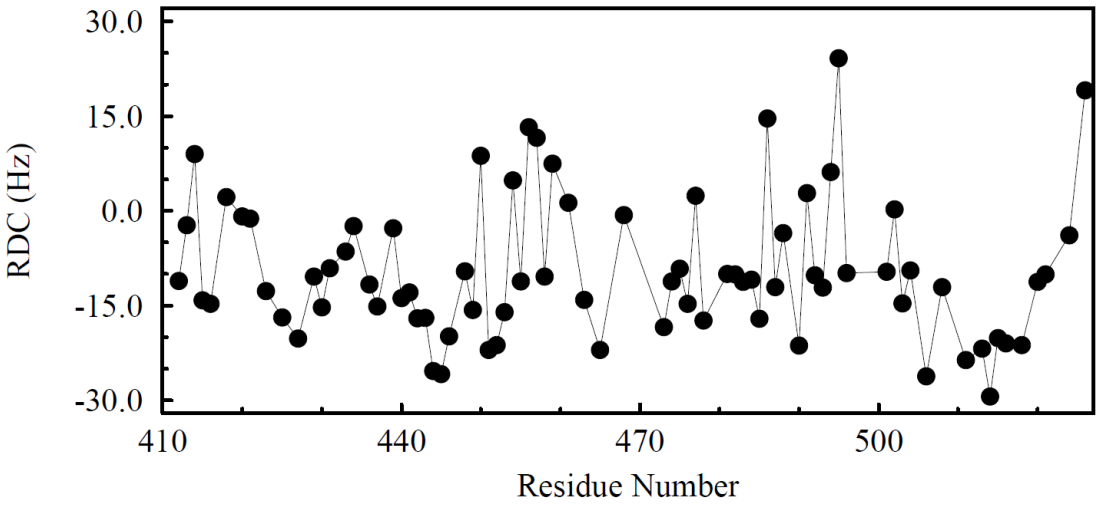


Figure 8.7 RDC measurements in hN-$1_{11\text{-}13}$. RDC values are spread between -30 Hz and +30 Hz.

8.3.2 Alignment tensors in individual domains

Fitting of an alignment tensor to EGF11 was started with the RDCs of 29 residues. Poor agreement was observed for residues 412, 413, 418 and 420, all of which had low heteronuclear NOE values below 0.45 (section 7.4), which demonstrates fast timescale (<~10 ns) backbone dynamics. Their orientation can therefore not be described by an alignment tensor fitted to the remaining overall rigid structure. Residue 434 was removed from the analysis as well since it contributed significantly to the overall error. The remaining 24 RDC values were used to fit the alignment tensor for EGF11, the results are summarised in Table 8.1. Errors in the alignment tensor were estimated using experimental errors of 2 Hz (a) and 4 Hz (b) in Monte Carlo simulations. A second method of error analysis was applied to estimate the significance of individual RDC couplings. Four residues were randomly eliminated and a standard deviation calculated for 100 runs (c). The small standard deviations in this latter method of error measurements indicate that the result of the fitting procedure is relatively insensitive to the specific group of RDC values that is used.

Table 8.1 RDC analysis of EGF11. RDC values showed a good fit using the crystal structure which is indicated by a low error and q-value. Alignment tensor values agree well in all measurements.

Domain	Nr. Residues	Error	Q	D_a (Hz)	R	Theta (°)	Phi (°)	Psi (°)
11a	24	3.8	0.13	18.6±0.5	0.25±0.02	19±1.5	164±4	132±5
11b	24	3.8	0.13	±1.1	±0.05	±3.0	±6.3	±9.5
11c	24	3.8	0.13	±0.2	±0.02	±1.1	±1.9	±2.2

a: Errors in alignment tensor parameters estimated using 500 Monte Carlo simulations with an experimental error of 2.0 Hz for RDC values
b: Errors in alignment tensor parameters estimated using 500 Monte Carlo simulations with an experimental error of 4.0 Hz for RDC values
c: Errors in alignment tensor parameters estimated using 100 calculations in which RDCs for four randomly selected residues were excluded.

As described above, four sets of Euler angles are compatible with a specific set of RDCs. The different possible angles are listed in Table 8.2.

Table 8.2 Four sets of Euler angles obtained for EGF11.

Orientation	Theta (°)	Phi (°)	Psi (°)
A	19	164	132
B	19	164	312
C	161	344	48
D	161	344	228

Figure 8.8 shows the four different orientations EGF11 can adopt according to rotations based on the Euler angles described in Table 8.2

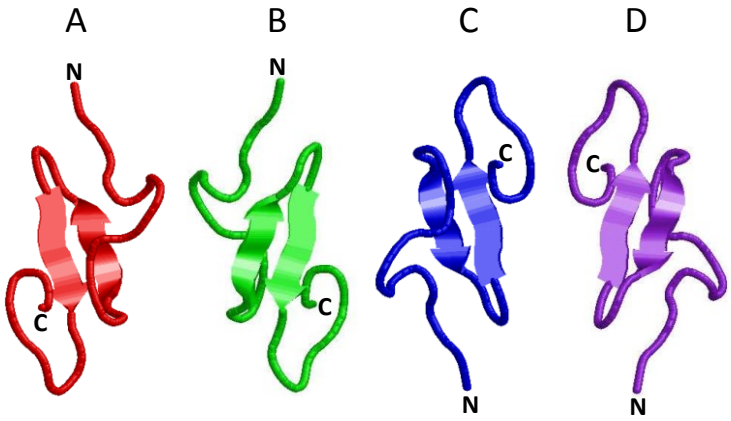


Figure 8.8 Domain orientations of EGF11. Rotations, according to the Euler angles described in Table 8.2, are applied. Note that rotations resulting along the short axis (C and D) are very unlikely because of the short length of the linker between EGF11 and EGF12.

Since EGF11 is covalently connected to EGF12, a rotation of the domain around the short axis by 180° is very unlikely. A rotation around the long axis, however, is possible but can be ruled out under certain circumstances, such as steric clashes or the availability of NOE contacts. This shows the requirement of a good starting model. If no model is available, this ambiguity can be resolved by the collection of additional sets of RDCs with different alignment tensors. However, this is difficult for EGF constructs because their elongated shape means that steric factors dominate the alignment.

Fitting an alignment tensor requires a good distribution of N-H bond vectors. The distribution of N-H vectors in the x/y-plane is shown in Figure 8.9 (left hand panel), the distribution along the z-axis is shown on the right hand panel.

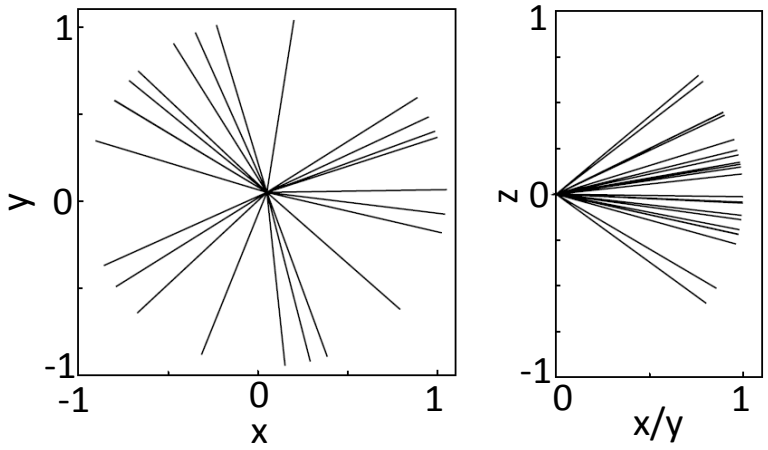


Figure 8.9 Distribution of N-H vectors of RDCs measured in EGF11. The distribution in the x-y direction is shown on the left, the distribution in the z-direction on the right. The overall distribution is good, vectors point in all direction and are not clustered.

The experimental RDC values of EGF11 are plotted against the calculated RDC values in Figure 8.10. Alignment tensors for EGF12 and EGF13 were fitted as demonstrated above for EGF11 (Table 8.3). Residue 461 in EGF12 and residues 491, 493, 504, 524 and 526 of EGF13 were removed from the analysis, since they contributed to high errors which could distort the alignment tensor values.

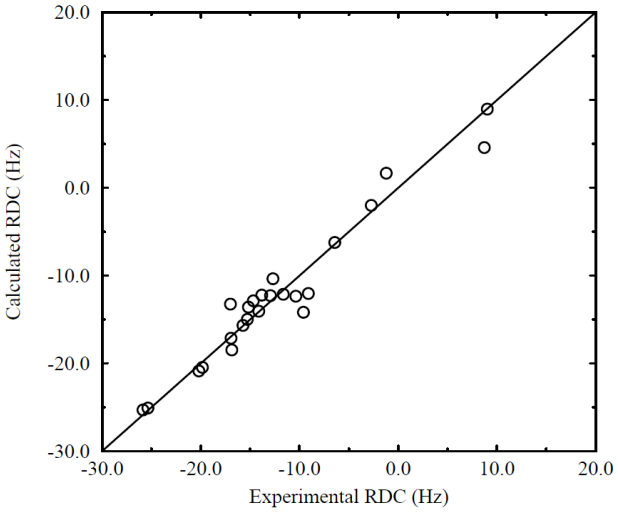


Figure 8.10 Fitting an alignment tensor to EGF11. The experimental values (horizontal axis) agree well with the calculated values (vertical axis).

The higher number of residues deleted when fitting EGF13 could arise from the instability of this domain described before (Hambleton et al., 2004).

Table 8.3 Alignment tensors fitted for EGF12 and EGF13 of the hN-$1_{11\text{-}13}$ construct.

Domain	Nr. Residues	Error	Q	D_a (Hz)	R	Theta (°)	Phi (°)	Psi (°)
12a	26	10.0	0.24	18.0±0.5	0.30±0.03	15.0±1.0	176.5±4.6	131.9±4.6
12b	26	10.0	0.24	±1.1	±0.06	±1.9	±9.4	±9.5
12c	26	10.0	0.24	±0.5	±0.02	±0.6	±4.2	±3.7
13a	18	8.1	0.16	21.0±0.7	0.26±0.03	24.1±1.1	176.5±2.5	147.6±4.4
13b	18	8.1	0.16	±1.3	±0.05	±2.2	±5.1	±12.4
13c	18	8.1	0.16	±0.6	±0.02	±1.6	±4.1	±7.5

a: Errors in alignment tensor parameters estimated using 500 Monte Carlo simulations with an experimental error of 2.0 Hz for RDC values.
b: Errors in alignment tensor parameters estimated using 500 Monte Carlo simulations with an experimental error of 4.0 Hz for RDC values.
c: Errors in alignment tensor parameters estimated using 100 calculations in which RDCs for four randomly selected residues were excluded.

This analysis shows that alignment tensors can be fitted to individual domains of the hN-$1_{11\text{-}13}$ construct with a reasonably low error value. The elimination of residues from the fitting process is necessary for the subsequent analysis of domain orientation, since the presence of a large error during the fit could lead to re-orientation based on the minimisation of a particular residue that has a large impact on the error.

8.3.3 Domain orientation

The crystal structure of hN-$1_{11\text{-}13}$ was rigid and elongated, suggesting that the whole constructs tumbles as a single rigid unit. This assumption is in line with the similarity of D_a and R values of the alignment tensors fitted to individual domains in EGF11-13. The values for D_a, for example, were roughly within the estimated error when an experimental error of 4 Hz was used (18.6±1.1 Hz for EGF11, 18.0±1.1 Hz for EGF12 and 21.0±1.3 Hz for EGF13). Differences in domain orientation between crystal structure and RDC fits can be estimated by comparing predicted Euler angles theta, phi and psi for individual domains. These were slightly different, suggesting that the three domains have a slightly different domain orientation in solution. Fits for the domain pairs EGF11-12, EGF12-13 and EGF11-EGF13 were carried out (Table 8.4).

First, alignment tensors were fitted using one set of D_a, R, theta, phi and psi (a). An improvement in the fits is obtained when the relative orientations of the domains are allowed to change (b) by using two sets of angles; this was more significant for the EGF12-EGF13 pairs, with the errors decreasing from 17.7 to 10.3 and from 15.2 to 6.5 for EGF11-EGF13.

Table 8.4 Alignment tensors in EGF11-EGF12, EGF12-EGF13 and EGF11-EGF13.

Domain	Nr. Residues	Error	Q	D_a (Hz)	R	Theta (°)	Phi (°)	Psi (°)
a:11-12	50	8.8	0.21	18.2±0.4	0.25±0.02	16.7±0.8	165.7±2.4	135.1±3.0
b:11-12	50	7.2	0.19	18.3±0.4	0.28±0.02	19.8±1.3	163.8±2.8	132.1±4.0
						14.9±0.9	175.6±4.6	132.7±4.5
a:12-13	44	17.7	0.27	19.7±0.4	0.29±0.02	18.0±0.7	176.3±2.2	140.1±2.9
b:12-13	44	10.3	0.21	19.2±0.4	0.30±0.02	15.0±0.9	177.3±4.4	132.0±4.3
						24.6±1.2	176.5±2.3	150.5±3.9
a:11-13	42	15.2	0.24	20.2±0.4	0.22±0.02	22.1±0.8	171.1±1.7	137.5±3.1
b:11-13	42	6.5	0.16	19.6±0.4	0.26±0.02	20.1±1.5	164.2±2.9	131.5±4.2
						24.4±1.1	176.9±2.2	149.2±3.9

a: Fitting was performed with one set of angles, and an experimental error of 2 Hz
b: Fitting was performed with the same D_a/R, but two independent sets of angles; the top line are angles for the 1st EGF and bottom line for 2nd EGF domain.

This leads to a new model of the solution structure of hN-$1_{11\text{-}13}$, in which the inter-domain twist and tilt angles can be refined. The error of the twist and tilt angles was estimated by

generating 100 sets of rotated coordinates using randomized theta, phi and psi angles within a Gaussian distribution with the width defined by the standard deviation obtained (Table 8.4). Table 8.5 shows the resulting twist and tilt angles of the domain pairs of the crystal and solution structure based on RDC measurements. The major beta-sheet was used as a reference in these measurements.

Table 8.5 Interdomain tilt and twist angles for EGF11-12 and EGF12-13 pairs. The error is estimated using a 2 Hz experimental error.

Domains	X-ray tilt (°)	X-ray twist (°)	RDC tilt (°)	RDC twist (°)
11-12	9.7	119.3	12.7 ± 1.5	132.8 ± 9.2
12-13	20.3	132.0	12.1 ± 1.3	148.6 ± 8.0

The domain pair EGF11-EGF12 shows good agreement when comparing the tilt angle based on the RDC data compared to the crystal structure. For the EGF12-EGF13 domain pair, the RDCs indicated a less tilted organisation, suggesting a more elongated structure. The twist angles were wider for both domains, although the error involved in the measurements was high. The change in inter-domain angles can be illustrated by rigid-body modelling based on the RDC values, using the computer program MODULE2 (Dosset et al., 2001); domains are rotated at the first position of an EGF domain (two residues after C6 in the preceding domain) (Figure 8.11).

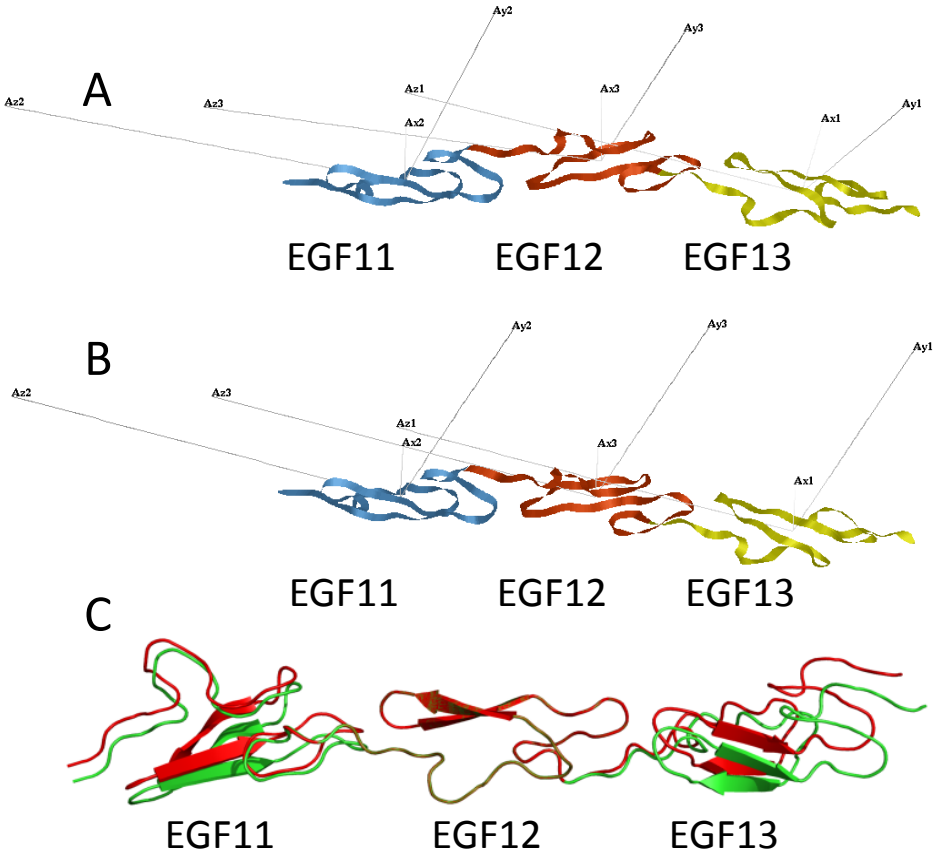


Figure 8.11 Rigid-body modelling using the software MODULE. An alignment tensor was fitted to individual domains (A). EGF11 is shown in blue, EGF12 in red and EGF13 in yellow. The eigenvectors are visualised. In (B), the alignment frame of individual domains is adjusted. In (C), EGF12 of the crystal structure (red) was overlayed to the modelled structure (green). Note the subtle changes in domain orientation.

The overall agreement for the triple hN-$1_{11\text{-}13}$ improved from 0.27 to 0.20 as visualised in Figure 8.12. Here, an alignment tensor fitted for the RDCs of EGF11 (black circles) was used to predict RDCs for EGF12 and EGF13 (open circles). The values for EGF12 and EGF13 agree much better after rigid-body modelling (B). Differences in inter-domain tilt and twist angles derived from the crystal structure and those measured by RDCs are relatively small, but not within one standard deviation of the measured error. However, the analysis of the crystal structure of hN-$1_{4\text{-}7}$ (section 6.5.1) indicated that in crystal structures too, differences of the same magnitude in tilt and twist angles can be observed.

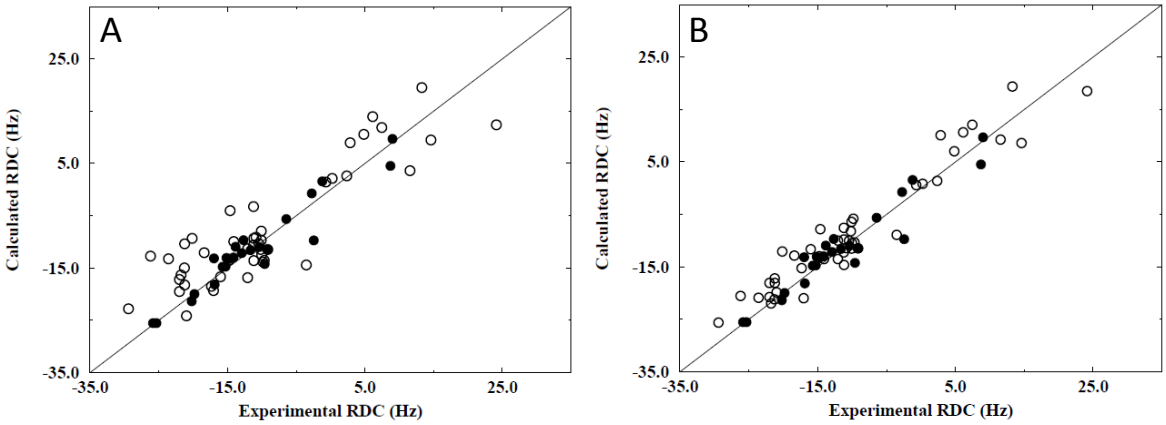


Figure 8.12 RDC data of hN-1_{11-13} on the crystal structure (A) and after rigid-body modelling (B). Filled, black circles indicate RDCs from EGF11 which were used to fit an alignment tensor. Open circles show the agreement of RDCs from EGF12 and EGF13. The agreement is better after rigid-body modelling (B), than before (A).

Several reasons are possible for an error in the RDC dataset that has not been taken into account yet. First, the dataset could have been too small and not sampled well enough the different possible NH orientations. However, Figure 8.9 indicated that vectors sampled a good range of orientations. A similar result was obtained for EGF12 and EGF13 (not shown). The solution structure of hN-1_{11-13} showed a rigid, rod-like orientation for EGF11 and EGF12, stabilised by Ca^{2+} (Hambleton et al., 2004), but the orientation of EGF11-12 was only approximated to 119±5° (twist) and 17±4° (tilt angle). The orientation of EGF13 could not be determined in that study, most likely due to µs-ms timescale motions in EGF13. Addition of the birA-tag at the C-terminus of EGF13 stabilised this domain and the crystal structure of hN-1_{11-13} showed a near rod-like conformation for all three domains. It is therefore possible that the rigidity of EGF13 in the crystal structure was overestimated or distorted by the non-native birA-tag.

8.3.4 Analysis of the hN-1_{11-13}12DG mutant

The introduction of the D469G substitution in EGF12 of the hN-1_{11-13} construct abrogated Ca^{2+} binding and ligand binding in FACS analysis with full-length extracellular human Delta-like-1 (Cordle et al., 2008b). In this section, the structural effect of this mutation is analysed by NMR,

since crystallisation of this construct was unsuccessful. RDCs were measured at a concentration of 2.5% peg/hexanol which resulted in a D_2O splitting of 9 Hz. The alignment medium was therefore slightly stronger than in the 2.0% peg/hexanol used for the hN-$1_{11\text{-}13}$ construct with a splitting of 7.1 Hz. RDCs are shown in Figure 8.13 and demonstrate a reduction in the magnitude of the RDCs, particularly for EGF11 and EGF12 when compared to hN-$1_{11\text{-}13}$ (Figure 8.7).

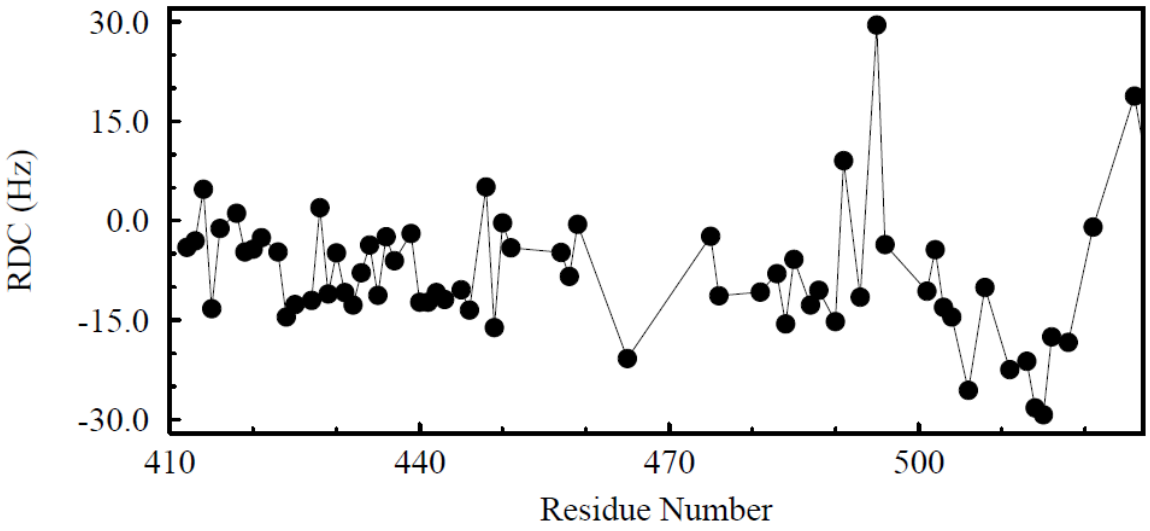


Figure 8.13 RDCs measured for the hN-$1_{11\text{-}13}$12DG mutant. The RDCs are significantly smaller than the RDCs observed for hN-$1_{11\text{-}13}$.

This suggests that hN-$1_{11\text{-}13}$12DG has a less elongated shape which correlates with the finding that the interface of EGF11-EGF12 is destabilised by the lack of Ca^{2+}. In ^{15}N-HSQC analysis, peaks belonging to EGF11 (D412-C440) and residues belonging to EGF13 including the inter-domain packing region after C5 in EGF12 (V489-Q526) were at identical positions in spectra of both the hN-$1_{11\text{-}13}$ and the hN-$1_{11\text{-}13}$12DG mutant constructs, suggesting structural identity of these regions (not shown). However, peaks after C5 in EGF11 (C440) belonging to the packing region with EGF12 were not at identical positions as were peaks for residues belonging to EGF12. The difference in chemical shift could be due to a structural rearrangement of both constructs or due to the change in chemical environment based on the presence or absence of Ca^{2+}. It was not possible to assign residues V453, N454 and E455 in between the sixth conserved cysteine of

EGF11 and the first conserved cysteine of EGF12, most likely due to line-broadening effects on the µs-ms timescale. Similarly, no additional strong peaks as expected for residues flexible on a very fast time-scale (<ns) were observed in the HSQC spectrum of hN-$1_{11\text{-}13}$12DG (not shown). Heteronuclear [^{1}H]-^{15}N NOE values for this construct were recorded to investigate whether or not the introduction of a Ca^{2+} binding mutation leads to flexible linker residues in between EGF11 and EGF12. Figure 8.14 shows a comparison of the data with the wildtype construct (section 7.4). Values for the hN-$1_{11\text{-}13}$ (top) were lower than 0.4 only at the N- and C-termini and for the loop region in between C1-C2 of EGF11, but higher than 0.6 for the residues in between C6 of EGF11 and C1 of EGF12. Similarly, the loop region and the termini of the hN-$1_{11\text{-}13}$12DG construct confirmed these regions to be flexible. Residues that were assigned in the region in between the sixth cysteine of EGF11 and the first cysteine of EGF12 were C449, E450, I451, D452 (Table 8.6).

Table 8.6 Heteronuclear [^{1}H]-^{15}N NOE values for the residues in between C6 of EGF11 and C1 of EGF12. Note that residues D452 and C456 show a lower hnNOE in the mutant.

	hN-$1_{11\text{-}13}$	hN-$1_{11\text{-}13}$12DG
C449	0.68 ± 0.10	0.73 ± 0.05
E450	0.75 ± 0.20	0.64 ± 0.08
I451	0.68 ± 0.09	0.70 ± 0.04
D452	0.85 ± 0.22	0.56 ± 0.04
V453	0.70 ± 0.13	-
N454	0.78 ± 0.12	-
E455	0.73 ± 0.10	-
C456	0.82 ± 0.10	0.43 ± 0.08

Whereas the heteronuclear NOE value was comparable to the wildtype for C449, E450 and I451, a lower value was observed for D452 and C456 with 0.56 ± 0.04 and 0.43 ± 0.08 and respectively. The combination of line broadening and the decrease in heteronuclear NOE value suggest flexibility in between EGF11 and EGF12 on different timescales. Interestingly, the peaks for residues of EGF11 in the hN-$1_{11\text{-}13}$12DG mutant were sharper, indicated by a lower error value when compared to EGF12-13 or the hN-$1_{11\text{-}13}$ wildtype.

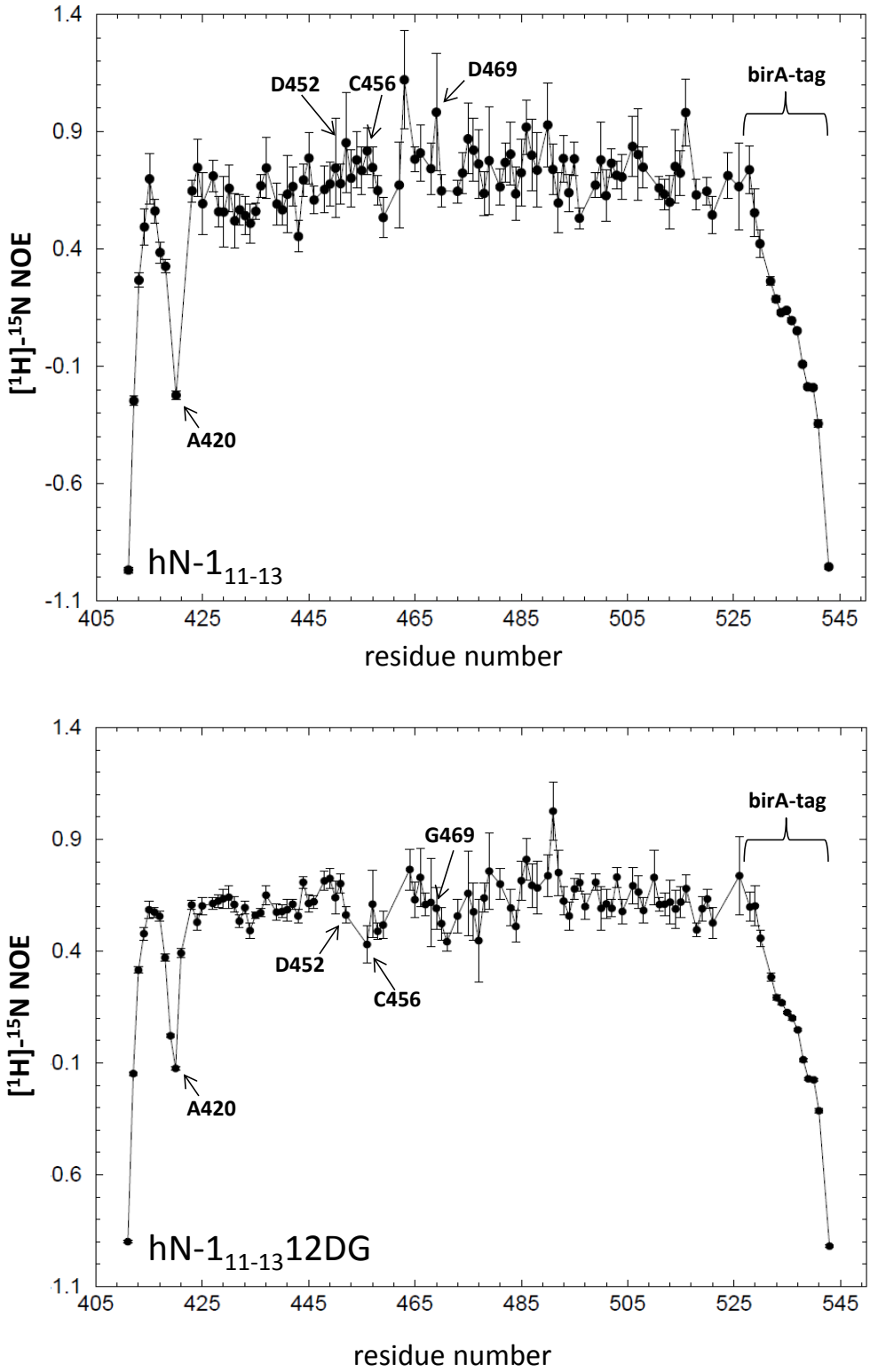


Figure 8.14 Heteronuclear [^{1}H]-^{15}N NOE data for hN-1$_{11\text{-}13}$(top) and hN-1$_{11\text{-}13}$12DG (bottom). Residues flexible on a fast timescale (<ns) were at the N- and C-termini and in the linker in between C1 and C2 of EGF11 as indicated by A420. The D469G mutation did not introduce flexibility on the fast time-scale at 469 and neighbouring residues. However, residues in between Cys6 of EGF11 and Cys1 of EGF12 showed a decrease in hnNOE for residues D452 and C456. V453, N454 and E555 were not assigned in the mutant construct, possibly due to line-broadening effects. The position of the flexible birA-tag at the C-terminus is indicated.

RDC analysis of EGF11 (D412–C440) using the coordinates of the hN-$1_{11\text{-}13}$ crystal structure, gave good agreement with a q-value of 0.15 using 19 residues (Table 8.7). The addition of the residues after C5 in EGF11 up to E450, 8 residues, did not change the q-value substantially suggesting that the inter-domain packing region is structurally not disturbed in the absence of Ca^{2+} binding in EGF12. However, fitting the alignment tensor resulted in slightly altered phi/psi angles. Given the high axial symmetry of this region, as demonstrated by the low R value, the phi and psi angles are not well defined and the difference could result from poor sampling because of a small number of RDC values. C-terminal residues of EGF12 and residues in EGF13 of the hN-$1_{11\text{-}13}$12DG dataset agreed well with the hN-$1_{11\text{-}13}$ structure and gave D_a, R and angles similar to those found in wildtype EGF12 and EGF13. In contrast, values for EGF12 did not agree well, suggesting a conformational change or flexibility in this region.

Table 8.7 Fitting domains of the hN-$1_{11\text{-}13}$12DG mutant. Whereas a low q-value suggests structural integrity for EGF11 and EGF13, the agreement of EGF12 was poor. Additional fitting is shown with identical residues from the hn-$1_{11\text{-}13}$ dataset to validate results. Monte Carlo simulations with an experimental error of 2 Hz for EGF12 and EGF13 were performed and 1 Hz for fittings of EGF11.

Domain	Res	Err	Q	D_a (Hz)	R	Theta(°)	Phi (°)	Psi (°)
11 (to 440)	19	1.9	0.15	12.2±0.3	0.11±0.03	4 ± 1	134 ± 18	24 ± 21
11 (to 450)	27	2.7	0.17	12.6±0.3	0.11±0.02	3 ± 1	83 ± 23	78 ± 24
12	12	9.4	0.34	10.4±0.9	0.22±0.07	23 ± 2	133 ± 8	98 ± 11
13 (490-529)	17	11.5	0.18	22.0±0.8	0.28±0.03	35 ± 1	170 ± 2	132 ± 4
13 (fr. 481)	21	16.4	0.24	19.8±0.6	0.28±0.02	33 ± 1	180 ± 2	131 ± 4
11and13(DG)	44	37.6	0.45	14.8±0.3	0.28±0.03	26 ± 2	174 ± 2	149 ± 4
11and13(wt)	44	16	0.25	19.5±0.4	0.25±0.02	23 ± 1	169 ± 2	141 ± 3
[m]11and13(DG)	44	20.6	0.33	15.7±0.4	0.21±0.03	4 ± 1	92 ± 2	68 ± 4
						33 ± 1	180 ± 2	140 ± 4
[ms]11and13(DG)	44	13.5	0.22	19.8±0.4	0.23±0.02	4 ± 1	114 ± 17	48 ± 27
						33 ± 2	179 ± 3	133 ± 4

[m] Fitting with two sets of angles
[ms] Fitting with two sets of angles, but scaling up the RDCs of EGF11 by 1.7

A large difference in D_a for EGF11 (D_a = 12.2) and EGF13 (D_a = 19.8 Hz or 22.0 Hz) indicated that the molecule doesn't tumble as a single rigid unit. Attempts to fit these RDC values with a single D_a and R and one or two sets of angles resulted in high q-values (0.45 and 0.33). The

measured residual dipolar couplings are scaled by the order parameter S (equation 5) of a particular NH vector. The decreased value of D_a for EGF11 compared to EGF13 could indicate that the order parameter for NH vectors in EGF11 has a value of ~0.6 relative to the order parameters of EGF13. If the RDC values in EGF11 are scaled up by ~1.7, to correct for the difference in S, then good agreement, with a q-value of 0.22 is obtained; this is similar to the q-value obtained for the wildtype when using identical residues. EGF13 of the hN-$1_{11\text{-}13}$12DG mutant, aligned to a similar extent as the hN-$1_{11\text{-}13}$ construct. However, a higher concentration of peg/hexanol was used for the hN-$1_{11\text{-}13}$12DG mutant, suggesting that for EGF12-13 too, the alignment is weaker in the hN-$1_{11\text{-}13}$12DG construct. The alignment tensor obtained using the scaled RDC values for EGF11 leads to an average structure with twist and tilt angles of 221° and 14° between EGF11 and EGF13 compared to values of 252° and 19° observed in the X-ray structure or 282° and 10° obtained using the wildtype RDC values. The change in these angles is likely to be the result of the loss in packing between EGF11 and EGF12.

Several possible types of motion could lead to the decrease in D_a for EGF-11 of the mutant. EGF11 could be less rigid as the packing to EGF12 is disturbed because of the lack of Ca^{2+} binding in EGF12. However, a low error value for RDC values in EGF11 and identical peak positions in the HSQC spectrum suggest that EGF11 does not experience significant destabilisation. Hence, it can be assumed that both ends of the hN-$1_{11\text{-}13}$12DG construct behave as rigid domains connected by a more flexible linking region.

Heteronuclear NOE experiments showed little evidence for fast timescale motions within EGF11, but indicated slightly higher mobility for residues at the beginning of EGF12 in addition to line broadening, most likely to µs-ms timescale motions. In a study (Tolman et al., 1997), it was shown that the local order parameter (S) can be related to a cone model, similar to the cone model introduced by Lipari and Szabo for sub-nanosecond motions observed in molecular dynamics simulations and NMR ^{15}N spin relaxation. In this model, the order parameter S relates

to the angle ψ that is the half angle of the cone. The smaller value of D_a for EGF11 than for EGF13 in the hN-$1_{11\text{-}13}$12DG mutant indicates that the centre of mobility is closer to the N-terminus than to the C-terminus. Here, the cone-model is used assuming that EGF12 and EGF13 are behaving as a rigid unit and EGF11 is rotating around these two domains. The angle of the cone within which this movement occurs can be approximated using equation (12, 13). Assuming an order parameter of 1 for EGF-13, the order parameter for EGF11 can be calculated by dividing the values of D_a of both domains (S = ~12.4/20.9 = ~0.6). Solving equation [13] using an order parameter of 0.6 results in an angle of approximately ~45 degrees through which EGF11 rotates around its average orientation, as shown in Figure 8.15.

$$D_{cone} = -SC_{int}\left[\frac{1}{2}\Delta\chi\left(3cos^2\theta - 1\right) + \frac{3}{4}\delta\chi sin^2\theta\cos(2\Phi)\right] \qquad (12)$$

where

$$S = \frac{1}{2}cos\Psi_{cone}(1 + cos\Psi_{cone}) \qquad (13)$$

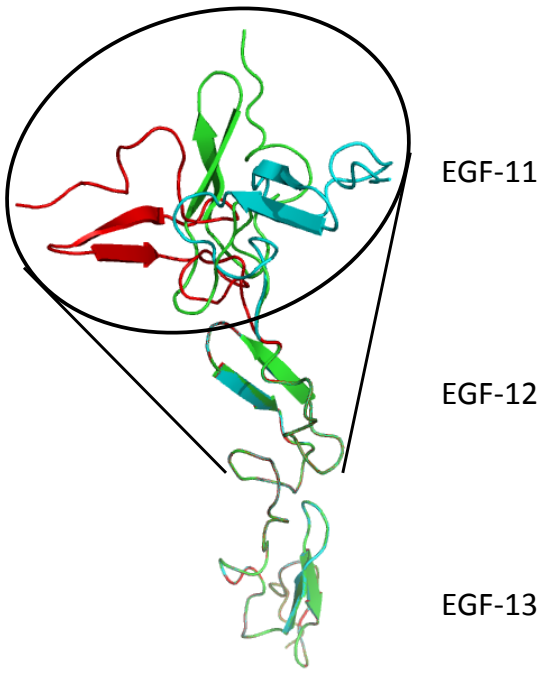


Figure 8.15 Motion of EGF11 with an angle of ~45° around domains EGF12-13 in the hN-$1_{11\text{-}13}$12DG mutant. Three structures of EGF-11 are shown to rotate within a cone.

These RDC data indicate that lack of Ca^{2+} binding in EGF12 through the D469G substitution leads to a loose interface between EGF11 and EGF12. Interestingly, the D469G mutation was shown to interfere with ligand binding (Cordle et al., 2008b) and was found in a case of squamous cell carcinoma (Wang et al., 2011), suggesting that the Ca^{2+} bound, rigid state is of physiological importance.

A physiological role for the Ca^{2+} unbound, non-rigid state remains to be identified. Under physiological conditions, the EGF11-EGF12 interface is likely to undergo a conformational exchange between rigid when Ca^{2+} is bound and non-rigid, when Ca^{2+} is not bound. Because of the high affinity of EGF12 for Ca^{2+}, the non-rigid state is only sparsely populated. Nevertheless, even a sparsely populated conformer could contribute to physiological effects, such as ligand binding or receptor activation. The transient loss of Ca^{2+} binding in EGF12 could allow the adoption of a different conformation, for example during receptor trafficking and receptor presentation. Once the alternate conformation is adopted, Ca^{2+} could bind again and lock the alternate conformation. Alternatively, the adoption of a bend in the EGF11-EGF12 interface as a result of transient loss of Ca^{2+} in EGF12 and additional long-range events such as dimerization or ligand binding, could result in the decrease of Ca^{2+} affinity and hence in an alternate stable structure.

Therefore, the role of Ca^{2+} binding in EGF12, its affinity and the role of the exchange between Ca^{2+} bound and unbound cannot be understood without a proper model of the cell surface organisation of the Notch receptor, ligand binding events and the activation of the receptor.

8.4 Analysis of the hN-1_{4-7} domain architecture

RDCs for the hN-1_{11-13} construct agreed with the crystal structure, for both individual EGF domains and with slightly different interdomain tilt and twist angles for the overall, rod-like

shape of the molecule. The hN-1_{4-7} crystal structure revealed a similar arrangement for domain pairs EGF4-5 and EGF6-7 with Ca^{2+} binding in EGF5 and EGF7 (section 6.5). However, a significantly different tilt angle was found in between EGF5 and non-calcium binding EGF6, that varied within the different crystal structures between 80-90°. Here, RDC analysis is used to test the domain orientations of hN-1_{4-7} in solution. RDCs were measured at 2.5% peg/hexanol with a D_2O splitting of 9.9 Hz (Figure 8.16). The RDCs were large, varying from -40 Hz to +40 Hz, a result comparable to the hN-1_{11-13} construct. However, the fact that hN-1_{4-7} has one domain more than hN-1_{11-13} and hN-1_{11-13}12DG, does not allow the prediction of the shape of hN-1_{4-7} on a simple comparison of alignment strength.

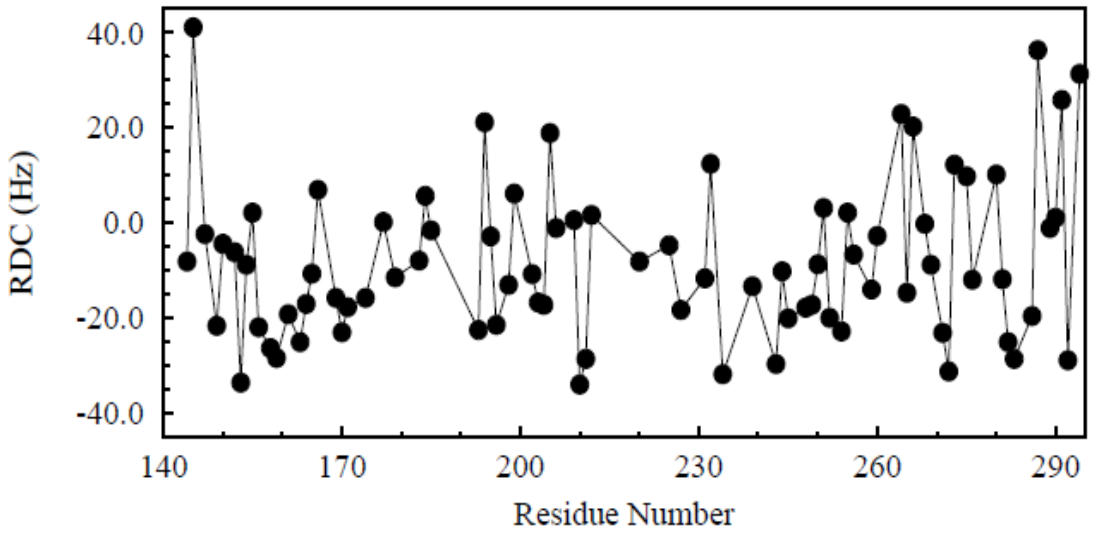


Figure 8.16 RDCs measured in the hN-1_{4-7} construct. RDC values range from -40 Hz to +40 Hz, showing strong alignment similar to the hN-1_{11-13} construct.

Alignment tensor parameters were fitted to individual EGF domains of the hN-1_{4-7} crystal structure molecule A in the P1211 structure (Table 8.8). For EGF4, 21 RDC values were measured. Two residues (144 and 161) were removed from the analysis due to low heteronuclear NOE values. For EGF5, 20 RDC values were measured. For EGF6, 18 RDC values were measured; the RDCs for residues 230 and 255 were removed from the analysis. For EGF7,

RDCs for 24 residues were measured. Poor agreement for residues 276, 280 and 282 was obtained and these RDCs were removed from the analysis.

Table 8.8 RDC analysis of individual domains in hN-1_{4-7}. Individual domains show a low q-value for all domains and similar alignment parameter values, except for phi and psi values of the pair EGF4-EGF5 when compared to EGF6-EGF7.

Dom	Res	Err	Q	D_a (Hz)	R	Theta (°)	Phi (°)	Psi (°)
4a	19	7.3	0.13	20.7±0.6	0.33±0.03	32±1	109±3	95±3
4b				±0.5	±0.02	±1	±3	±3
5a	20	12.3	0.22	20.3±0.8	0.35±0.04	25±2	109±3	84±4
5b				±1.0	±0.04	±2	±3	±5
6a	16	10.6	0.19	21.3±0.7	0.22±0.03	37±2	152±3	45±4
6b				±1.0	±0.04	±3	±3	±3
7a	21	7.5	0.14	19.1±0.6	0.41±0.03	30±1	164±3	47±4
7b				±0.8	±0.03	±1	±3	±5

a: the error was estimated in Monte Carlo simulations with 2 Hz experimental error.
b: the error was estimated using 100 calculations with 4 randomly deleted residues.

Fitting of 19 residues to EGF4 resulted in a good agreement of err = 7.3 and q = 0.13, similar to the other domains with q-values below 0.23. All values for D_a were around ~19-21 Hz, indicating that all EGF domains align to a similar extent suggesting that the whole molecule tumbles as a single rigid unit. The values of R vary slightly, and range between 0.22 for EGF6 to 0.41 in EGF7. The orientation of the domains with respect to the alignment frame was similar for domain pairs EGF4-5 and for EGF6-7, with the Euler angle theta being around 30° for all domains. The largest deviation of angle was observed for the phi and psi angles which are different for EGF4 and EGF5 when compared to the domains EGF6 and EGF7. This suggests a difference in the EGF5-EGF6 interface in solution.

Domain pairs were fitted in the same manner as for the analysis of hN-1_{11-13} (Table 8.9). First, fittings were performed assuming the same alignment tensor parameters. The angles theta, phi and psi were subsequently allowed to adopt independent values for individual domains. The agreement was good for domain pairs EGF4-5 with a q-value of 0.21 that is well in between the q-value of the individual domains (0.13 for EGF4 and 0.22 for EGF5). The agreement

is also acceptable for EGF6-7 with a q-value of 0.21. However, the error for the domain pair EGF5-6 increased significantly to q=0.38. Allowing the alignment tensor to adopt two different angles improves the fit for the EGF5-6 pair drastically to q=0.21. Similarly, the overall structure, when allowing two different sets of angles for the EGF4-5 pair and the EGF6-7 improved the fit from q=0.33 to q=0.21.

Table 8.9 Alignment tensor values for domain pairs in hN-1_{4-7} and the whole molecule. Fitting an alignment tensor with two sets of angles improves the fit of the domain pair EGF5-EGF6 as well as when fitting EGF4-5 against EGF6-7.

Domains	Residues	Error	Q	D_a (Hz)	R	Theta (°)	Phi (°)	Psi (°)
4-5a	41	13.8	0.21	20.9±0.4	0.31±0.02	29±1	110±2	93±3
4-5ma	41	11.5	0.19	21.1±0.4	0.30±0.02	32±1	109±3	94±4
4-5mb				±0.5	±0.03	±1	±3	±2
4-5m2a						26±1	110±3	83±4
4-5m2b						±2	±3	±4
5-6a	36	39.9	0.38	19.7±0.5	0.27±0.03	29±1	128±2	75±3
5-6ma	36	12.3	0.21	21.0±0.5	0.28±0.02	26±1	111±2	83±3
5-6mb				±0.7	±0.03	±2	±2	±5
5-6m2a						37±2	152±3	45±4
5-6m2b						±3	±3	±3
6-7a	37	16.1	0.21	19.8±0.4	0.36±0.02	32±1	162±2	46±3
6-7ma	37	10.8	0.17	20.4±0.4	0.32±0.02	37±2	153±3	45±4
6-7mb				±0.6	±0.02	±3	±4	±3
6-7m2a						30±1	165 ±3	44±4
6-7m2b						±1	±3	±4
45-67a	78	36.7	0.33	19.2±0.3	0.34±0.02	28±1	128±2	80±2
45-67ma	78	15.1	0.21	20.3±0.3	0.33±0.02	29±1	109±2	94±3
45-67mb				±0.2	±0.01	±1	±1	±2
45-67m2a						31±1	162±2	45±3
45-67m2b						±1	±2	±2

m: alignment tensors were fitted using two sets of angles
a: error analysis using 500 Monte Carlo simulations with an experimental error of 2.0 Hz for the RDC values.
b: error analysis using 100 calculations in which 4 residues were removed randomly
2: second set of angles

Analysing the twist and tilt angles from the reorientation of domains (Table 8.10), showed that the EGF5-EGF6 interface has a tilt angle of 90.3° in the crystal structure, but only 70±3° after rigid body modelling using RDCs for EGF5-6 or to 65° when using RDCs for EGF4-5 and EGF6-7. Figure 8.17 shows a plot of the complete RDC dataset based on an alignment tensor for EGF4-5.

Table 8.10 Interdomain tilt/twist angles for domain pairs in hN-1_{4-7} measured with mod2. Tilt and twist angles were analysed before and after rigid-body modelling using the RDC dataset. Note the decrease in tilt angle of the EGF5-EGF6 interface.

	X-ray		From RDCs	
Domains	**Tilt (°)**	**Twist (°)**	**Tilt (°)**	**Twist (°)**
	β-sheet /Cys	β-sheet /Cys	β-sheet /Cys	β-sheet /Cys
4-5a	34.0/35.0	179.2/175.2	43.5±2.2/44.4±2.2	185.6±5.2/181.4±5.2
4-5b			43.5±4.0/44.3±4.0	185.6±9.0/181.5±9.0
5-6a	90.3/88.0	112.6/118.4	69.6±2.1/67.9±1.8	120.2±6.5/124.3±6.5
5-6b			69.8±4.2/68.0±3.7	122.2±11.8/126.3±11.9
6-7a	38.8/42.2	152.6/159.6	42.9±2.0/45.2±1.9	152.4±4.2/159.8 4.1
6-7b			44.0±4.3/46.3±4.1	151.1±8.6/158.5±8.4
45-67a	90.3/88.0	112.6/118.4	65.0±1.7/62.3±1.5	120.9±4.4/124.4±4.4
45-67b			65.5±3.1/62.8±2.8	122.8±8.7/126.4±8.7

a: Standard deviation based on an experimental error in RDC measurement of 2.0 Hz
b: Standard deviation based on an experimental error in RDC measurement of 4.0 Hz

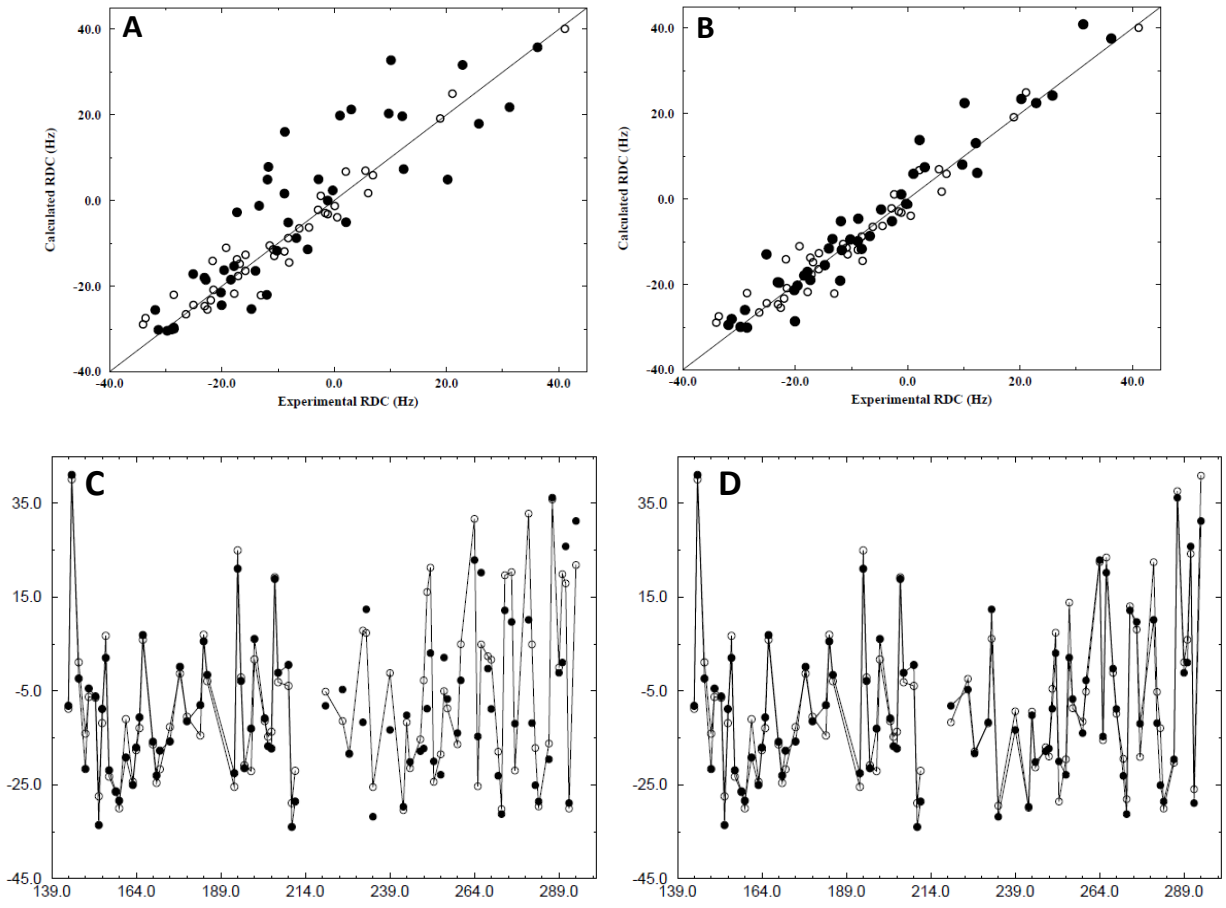


Figure 8.17 Plot of all RDCs of hN-1_{4-7} using the crystal structure (A and C) and after rigid-body modelling (B and D). In A, open white circles represent the RDC values for EGF4-5 (139-212) that were used to fit the alignment tensor. The RDC values for EGF6-7 (213-294) are shown as black circles and were calculated using the alignment tensor for EGF4-5. (B) The RDC values are calculated using the rigid-body modelled structure. In C and D, experimental (black circles) and calculated RDC values (open white circles) are plotted versus sequence.

With this alignment tensor, the RDCs for EGF6-7 were calculated and the plots on the right show the significant improvement in the fit following rigid-body modelling.

Figure 8.18 shows the Eigenvectors of EGF4-7 using MODULE. Alignment tensors were fitted to individual domains (A) and the eigenvectors of individual domains aligned (B). Note the change in the tilt angle between EGF5 and EGF6.

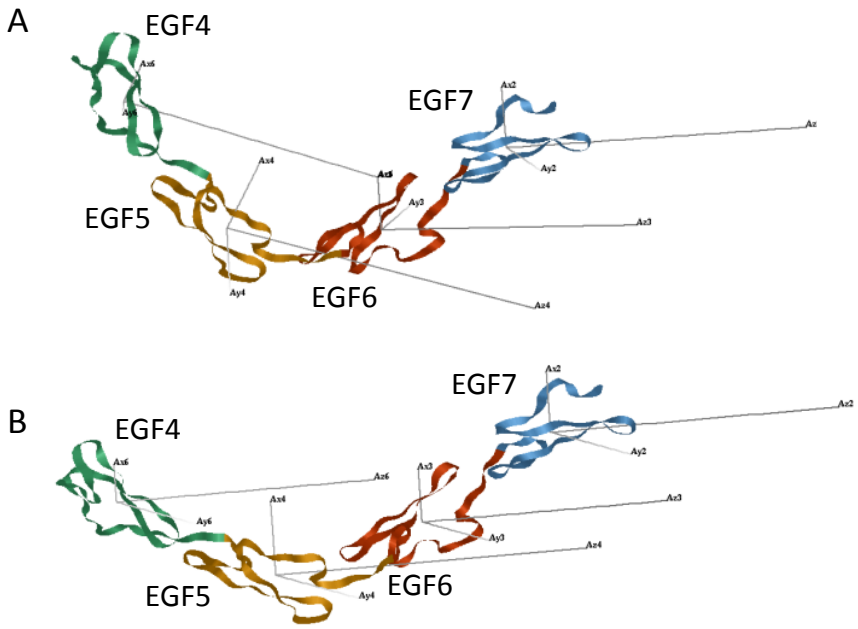


Figure 8.18 Crystal structure of hN-1_{4-7} (A) and after refinement using MODULE (B). Note that the angles in between 4-5 and 6-7 remained the same whereas the tilt angle in between 5-6 decreased.

The reason for the difference in angle is unclear. The interdomain packing within the EGF5-6 domain pair is formed in between A208 and V239. Usually, the packing residue of the N-terminal domain is an aromatic residue that because of its larger hydrophobic side-chain system could potentially form a tighter interface than alanine. Whereas similar strength of alignment is often referred to indicate domain rigidity, D_a values for the domain pairs EGF4-5 and EGF6-7 could be similar because of the symmetrical shape of the construct, in which EGF4-5 and EGF6-7 align independently, but with similar tensors. However, if there were large, fast timescale motions, it wouldn't be possible to align all domains to a common alignment tensor, and the RDCs would be

scaled down, similar to the hN-$1_{11\text{-}13}$12DG construct. Similarly, the analysis by heteronuclear NOE of the hN-$1_{4\text{-}7}$ construct does not give any hint of any fast-timescale motions in between EGF5 and EGF6 (Figure 8.19). Two more datasets of the hN-$1_{4\text{-}7}$ construct were analysed. The pH dependence was excluded by increasing the pH from 6.1 to 7.5. Another alignment medium with a lower percentage of peg/hexanol (2%) was tested as well. None of these gave a significantly different result. It is possible therefore that the crystal structure does not represent the solution structure of EGF4-7 well.

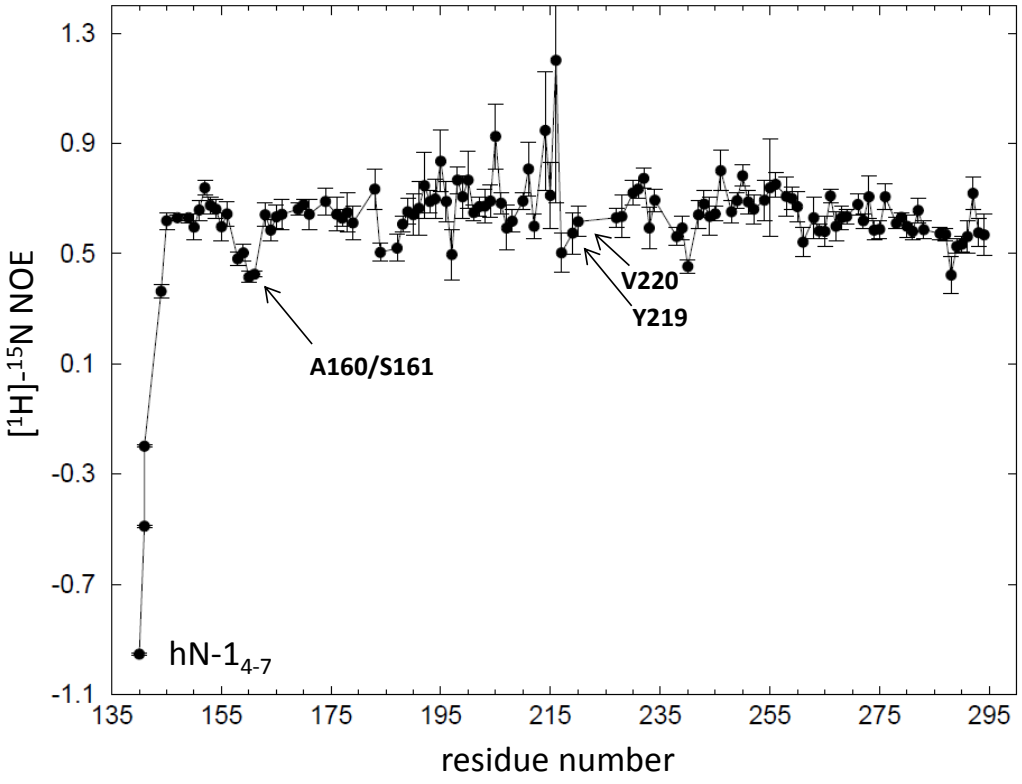


Figure 8.19 Heteronuclear [^{1}H]-^{15}N NOE data for hN-$1_{4\text{-}7}$. Residues flexible on a fast timescale (<ns) were at the N-terminus of EGF4 agreeing with the absence of those residues in the electron density map (chapter 6). Similarly, the poorly defined region at the end of the turn of the β-sheet that was not well defined in the crystal structure, showed a slight reduction for A160/S161. Y219/V220, the residues that are predicted to pack and to maintain the bent shape of EGF5-6 don't show mobility on a fast timescale but are also not very rigid.

Figure 8.20 shows a spacefill view of the structure of hN-$1_{4\text{-}7}$, with a clearly less defined EGF5-6 interface when compared to the EGF4-5 and EGF6-7.

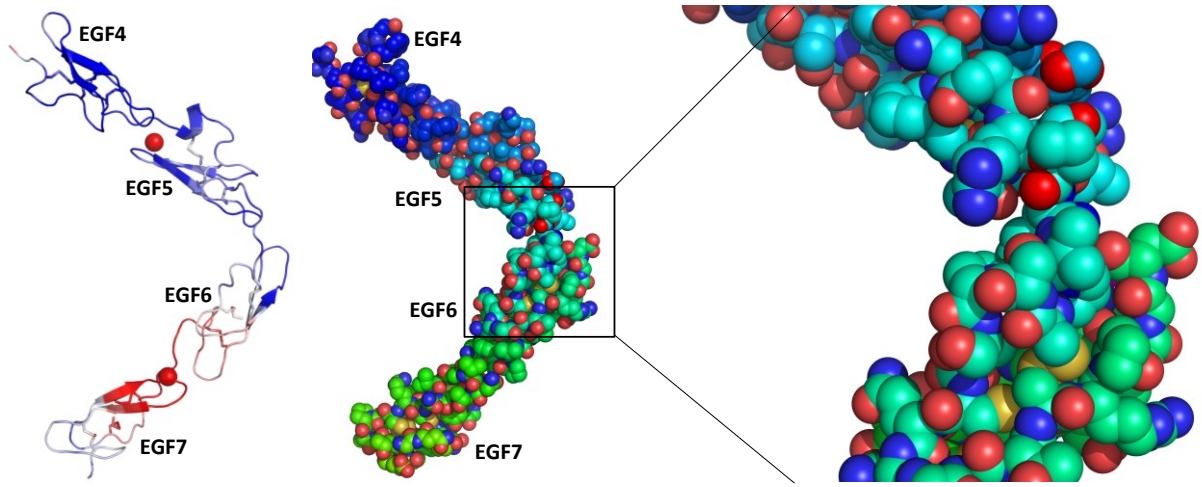


Figure 8.20 Temperature factors (left) and a space-fill (middle and right) representation of (the C121 crystal structure). The B-factors (temperature) were indicated ranging from 20 (blue) to 50 (red) showing higher temperature factors for the EGF6-7 rather than the EGF5-6 interface. A space-fill representation indicates that the interface between EGF5-6 is not extensive.

8.5 Analysis of the EGF7-8 domain architecture

The domain arrangement of hN-1_{7-8} was studied in the construct hN-1_{7-9}. Whereas 2D NOESY and Ca^{2+} chelation experiments indicated calcium binding in all three cbEGF domains, with a K_d of 19 ± 3 µM for EGF9 (Figure 4.17 and Figure 4.18, section 4.5), HSQC peaks for EGF9 showed line broadening, similar to EGF13 in the hN-1_{11-13} construct (Hambleton et al., 2004). The addition of residues at the C-terminus, in the form of the birA tag, increased the stability and folding of EGF13 (Cordle et al., 2008a, Hambleton et al., 2004). However, the addition of a birA tag to the hN-1_{7-9} construct had no effect on HSQC peak intensities in EGF9 (not shown). The instability of EGF9, may have prevented the hN-1_{7-9} construct from crystallising. Here, RDCs are used to model the interface of EGF7-8 using available crystal structures of other regions with similar structural properties (Figure 8.21). Figure 8.21 shows schematic diagrams of EGF7-8 (left) and of two domain pairs with similar organisation, EGF4-5 in hN-1_{4-7} and EGF11-12 of hN-1_{11-13}. Note that the C1-C2 loop lengths in EGF7-8 are the same as in EGF 4-5 but differ from those found in EGF11-12.

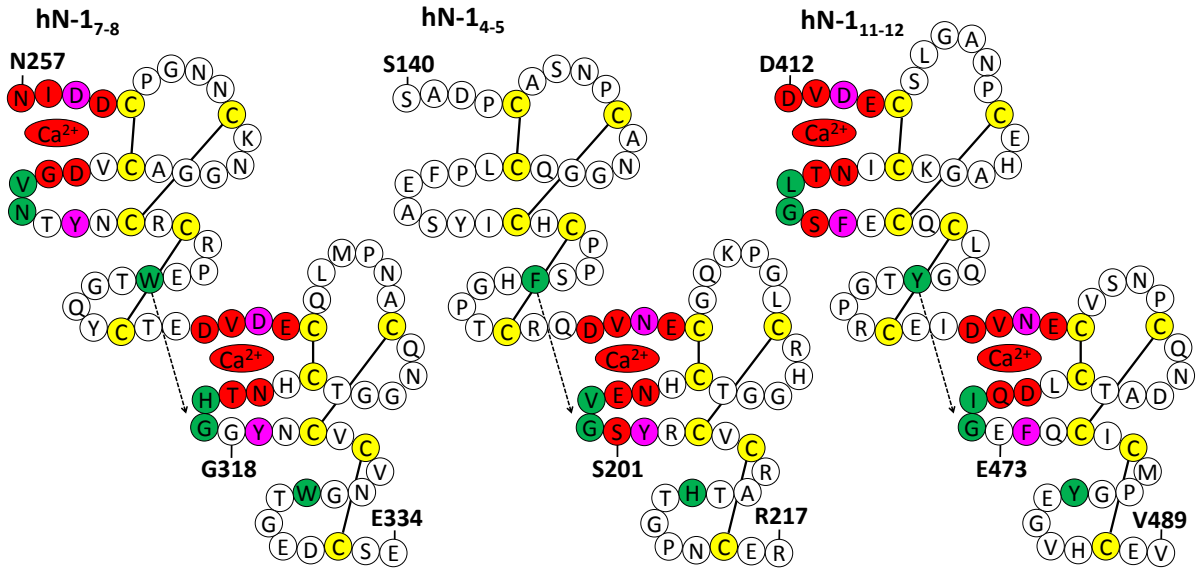


Figure 8.21 Schematic diagrams of hN-1$_{7-8}$ (left hand panel), as well as hN-1$_{4-5}$ (middle panel) and hN-1$_{11-12}$ (right hand panel). Note that in the crystal structures the twist angle between EGF11 and EGF12 is 119° and the angle between EGF4 and EGF5 is 179° (chapter 6.4, table 6.1), which was attributed to the donation of a carbonyl ligand by S201 in EGF5, but not by E473 in EGF12. It is unclear whether the corresponding residue in EGF8, G318 donates this ligand or not. In the schematic diagram, calcium and ligands are depicted in red, interdomain packing residues in green, cysteines in yellow and conserved residues for intradomain stability in purple.

RDCs for EGF7-8 of the hN-1$_{7-9}$ construct were collected at 2.5% peg/hexanol (D_2O splitting 9.1 Hz) and are presented in Figure 8.22.

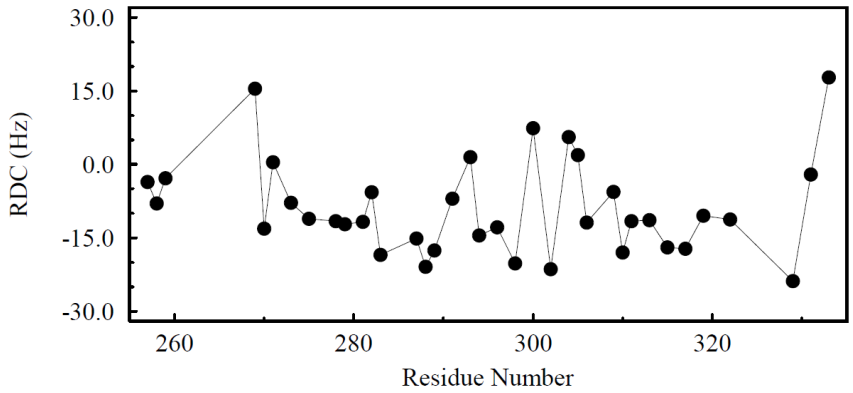


Figure 8.22 RDCs measured for EGF7 (residues 257 to 293) and EGF8 (residues 294 to 333) in the hN-1$_{7-9}$ construct.

RDCs collected for EGF7 were first fitted using the coordinates of EGF7 in the crystal structure of hN-1$_{4-7}$ (Table 8.11, 7$_{crys}$). Residues 257, 258 and 293 were removed from the

analysis. These residues are located at the N- and C-termini of EGF7 and differences in their conformation might be expected as a result of the different context in which EGF7 is placed in hN-1_{4-7} and hN-1_{7-9}. A good fit is obtained with the remaining 16 residues of EGF7 indicating that the overall structure of EGF7 is the same in the hN-1_{4-7} and hN-1_{7-9} constructs.

Table 8.11 RDC analysis of EGF7-8 using the crystal structure coordinates of hN-1_{4-7} and a homology model based on the hN-1_{11-13} crystal structure.

Domain	Model	Residues	Err	Q	D_a (Hz)	R	Theta (°)	Phi (°)	Psi (°)
7	**7_{crys}**	16	5.4	0.18	18.3 ± 0.8[a]	0.27 ± 0.04	79 ± 2	35 ± 2	116 ± 4
7	**4_{crys}**	15	5.2	0.18	16.2 ± 0.7	0.25 ± 0.04	59 ± 2	75 ± 2	69 ± 4
7	**11_{hom}**	14	8.3	0.22	16.3 ± 0.8	0.23 ± 0.03	84 ± 3	3 ± 3	126 ± 5
8	**5_{crys}**	16	8.4	0.19	17.5 ± 0.7	0.24 ± 0.03	47 ± 2	41 ± 3	100 ± 4
8	**12_{hom}**	15	6.7	0.18	19.6 ± 0.9	0.24 ± 0.04	83 ± 2	13 ± 2	6 ± 4

[a]In sections 8.3.2, it was shown that similar errors in the alignment tensor variables were obtained by randomly removing four residues from the fit as compared to 500 Monte Carlo simulations with 2 Hz experimental error. Therefore, the error for these fits is based only on the Monte Carlo simulations with an experimental error of 2 Hz.
$_{crys}$ RDCs were fitted on the crystal structure coordinates
$_{mod}$ RDCs were fitted on the coordinates of a homology model

RDCs collected for EGF7 and EGF8 were next fitted using the coordinates of EGF4 and EGF5 within the crystal structure of hN-1_{4-7} (Table 8.11, 4_{crys} and 5_{crys}). Residues 257, 258 and 259 were removed from the fit, since they correspond to the N-terminal region in EGF4 that was not defined in the crystal structure (Figure 7.21, section 7.5); residue 275 was also removed due to poor agreement in the fits. The remaining RDCs all agreed well with the crystal structure of EGF4 (err = 5.2, q = 0.18); this is in line with the finding that torsion angles of EGF domains and cbEGF domains are not significantly different (Figure 6.15, section 6.5.1). In EGF8, good agreement with the crystal structure of EGF5 (err = 8.4, q = 0.19) was obtained by removing residues 304 and 305 from the fit; these residues belong to the flexible loop region between C1 and C2 (Figure 7.20, section 7.4). The RDCs for EGF7 and EGF8 were also fitted using EGF11 and EGF12 in the hN-1_{11-13} crystal structure. Because of differences in the lengths of the C1-C2 loops (Figure 8.21), a homology model was built using the MODELLER software (Eswar et al., 2007). For EGF7, good agreement (err = 8.3, q = 0.22) was obtained using the homology model by removing the N-

terminal residues, 457-459, and residue 271 from the fit. The RDCs for EGF8 were also fitted using the homology model; good agreement (err = 6.7, q = 0.18) was obtained by removing residues 302 and 304-306 (loop region) from the fit. These results show that RDCs can be fitted to crystal structures of different domains with only limited sequence identity.

The D_a values for EGF7 varied from 16.2 ± 0.7 Hz to 18.3 ± 0.7 Hz while the D_a values for EGF8 varied from 17.5 ± 0.7 Hz to 19.6 ± 0.9 Hz, indicating some lack of precision when fitting a small number of RDCs to an imperfect model. However, the difference between 16 Hz and 19 Hz is significantly smaller than the difference in D_a observed for EGF11 and EGF13 of the hN-$1_{11\text{-}13}$12DG mutant (~12 Hz for EGF11 and ~20-22 Hz for EGF13), therefore suggesting that EGF7 and EGF8 are tumbling as a single rigid unit. The values for the angles theta, phi and psi were different according to the model used; these represent the Euler angles necessary for each model to align A_{zz}, A_{xx} and A_{yy}, the principle components of the alignment tensor, with the Cartesian z, x and y axes. The values for R show good agreement between EGF7 and EGF8, in agreement with a rigidly rotating domain pair. Alignment tensors were fitted for the domain pair EGF7-8 using both the crystal structure of EGF4-5 in hN-$1_{4\text{-}7}$ and the homology model based on EGF11-12 (Table 8.12).

Table 8.12 Measurement of tilt and twist angles for the EGF7-8 interface.

Dom	Mod	Res	Err	Q	D_a (Hz)	R	Twist (°)	Tilt (°)
7-8	**45_{mod}**	30	14.2	0.27	17.2±0.5	0.23±0.03	174.4	35.5
7-8	**$45_{mod\text{-}arr}$**	32	8.4	0.21	17.0±0.5	0.21±0.03	178±7	31±3
7-8	**1112_{mod}**	29	18.4	0.31	17.1±0.5	0.13±0.03	112.1	22.3
7-8	**$1112_{mod\text{-}arr}$**	31	10.8	0.24	17.6±0.5	0.26±0.03	177±7	28±4

$_{mod}$ tilt and twist angles of the models used
$_{mod\text{-}arr}$ tilt and twist angles after remodelling the position of EGF7-8 with respect to each other.

As in previous sections, fitting was performed using a single set of angles, to test whether the relative orientation of domains in the models was correct, and using two sets of angles, to determine the best relative orientation of the domains. For both models, the error values

decreased for the fit with 2 sets of angles. However, the decrease was more significant for the homology model based on EGF11-12, where the error decreased from 18.4 to 10.8, as compared to a decrease of 14.2 to 8.4 for the crystal structure of EGF4-5. The resulting twist angles of around 180° for both models are very similar to the twist angles observed for the EGF4-5 (179.2°) and EGF22-23 (171.4°) pairs (Table 8.12). Similarly, the tilt angle is around 30° for both models, which is in line with the tilt angles measured for EGF4-5 (34°), and a bit smaller than the tilt angle of EGF22-23 (42.6°). Errors for twist and tilt angles were estimated as described (section 8.3.3).

Figure 8.23 shows the correlation between experimental and calculated RDCs based on the homology model of EGF11-12, before (A) and after (B) rigid-body modelling. An alignment tensor was fitted to EGF7 with which RDCs were predicted for both EGF7 (filled circles) and EGF8 (open circles). The agreement is much better after rigid-body modelling (B).

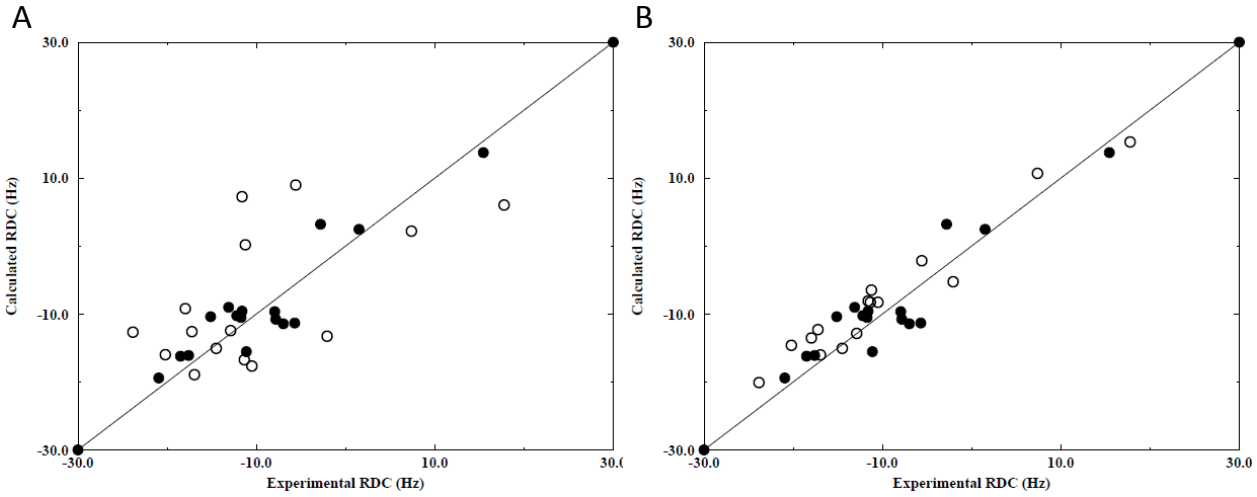


Figure 8.23 RDC fitting for EGF7-8 using the homology model of EGF11-12 before (A) and after (B) rigid body modelling. In both cases, the alignment tensor was fitted to EGF7 only; predicted RDCs for EGF8 (open) circles do not agree well with the initial model (A), but do agree well after rigid body modelling (B).

Figure 8.24 shows the resulting interface for EGF7-8, which was obtained using rigid-body modelling using the graphical interface MODULE2 (Dosset et al., 2001). Similar to the interface of

EGF4-5 (section 6.5.4), the aromatic packing residue W287 packs with H316 in a different orientation when compared to the EGF11-EGF12 interface. The similarity in tilt and twist angles between EGF7-8 and EGF4-5/EGF22-23 suggests that this domain may have the additional backbone carbonyl Ca^{2+} ligand, which was not clear from the analysis of $^{15}N,H^N$ and C' chemical shifts (section 7.5). Interestingly, the corresponding donating residue 318 is a glycine, which is similar to EGF23 with glycine 890 (section 6.6.2), whereas previous work suggested that only serines or threonines can donate the extra backbone carbonyl ligand because of their hydroxyl group which forms other interactions (Lee et al., 2004).

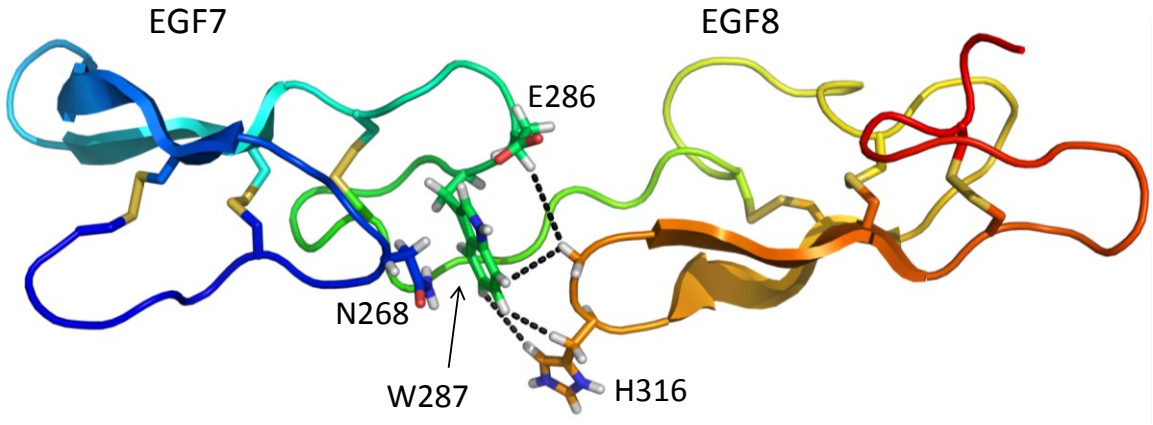


Figure 8.24 Model of hN-1_{7-8} based on a homology model of hN-1_{4-5} and after domain rearrangement. Note that the packing is very different when compared to EGF11-12 (section 6.5.4), but very similar to that of EGF4-5. The packing of the conserved aromatic residue W287 (EGF7) with H316 (EGF8) is indicated.

8.6 Analysis of the hN-1_{8-11} domain architecture

Similar to the experience with hN-1_{7-9}, crystallography was not successful for the hN-1_{8-11} construct. In this analysis, a combination of crystallographic coordinates from the hN-1_{11-13} and hN-1_{21-23} structures were used to fit a set of RDCs collected for hN-1_{8-11} at around 2.7% peg/hexanol (D_2O splitting of 10.3 Hz) (Figure 8.25). Models were constructed for the EGF8-EGF9 and EGF10-EGF11 interfaces and not for the EGF9-EGF10 interface, since cbEGF-cbEGF and EGF-cbEGF interfaces have shown less variation in tilt and twist angles as compared to vbEGF-cbEGF

interfaces; their structural characteristics can therefore be better anticipated (Table 6.1, section 6.4).

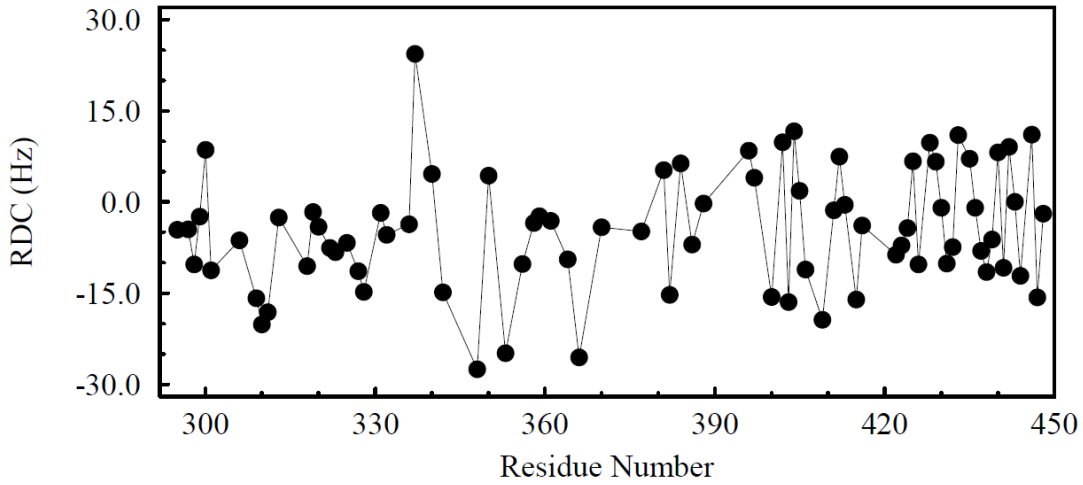


Figure 8.25 RDCs collected for the hN-$1_{8\text{-}11}$ construct are plotted along the sequence of hN-$1_{8\text{-}11}$.

The EGF8-EGF9 pair was modelled using the crystal coordinates of EGF11-EGF12 (Figure 8.26). EGF8 and EGF11 both have 6 residues in the C1-C2 loop. EFG9 and EGF12 both have 4 residues in the C1-C2 loop. In addition, chemical shift analysis described in section 6.4 suggests that S356 in EGF9 does not provide a carbonyl ligand for Ca^{2+} binding; this is also the case for EGF12. Results of fitting alignment tensors to the individual domains are presented with error estimates based on a Monte Carlo simulation with 2 Hz experimental error (Table 8.13). Residues 302, 304 and 305 of the flexible loop between C1 and C2 in EGF8 were removed. Good agreement for EGF8 was obtained after the further removal of residues 310-312, 326 and 330, resulting in an error of 4.0 when fitting 19 residues. For the fit of EGF9, residues 343, 363 and 365 were removed, and the fit with 14 residues had an error value of 3.5. The D_a value for EGF9 (17.9 Hz) was higher than for EGF8 (11.7 Hz), suggesting that this interface may not tumble as a rigid unit, although this difference is smaller than in the case of the hN-$1_{11\text{-}13}$12DG mutant (EGF11 12 Hz, EGF13 20-22 Hz).

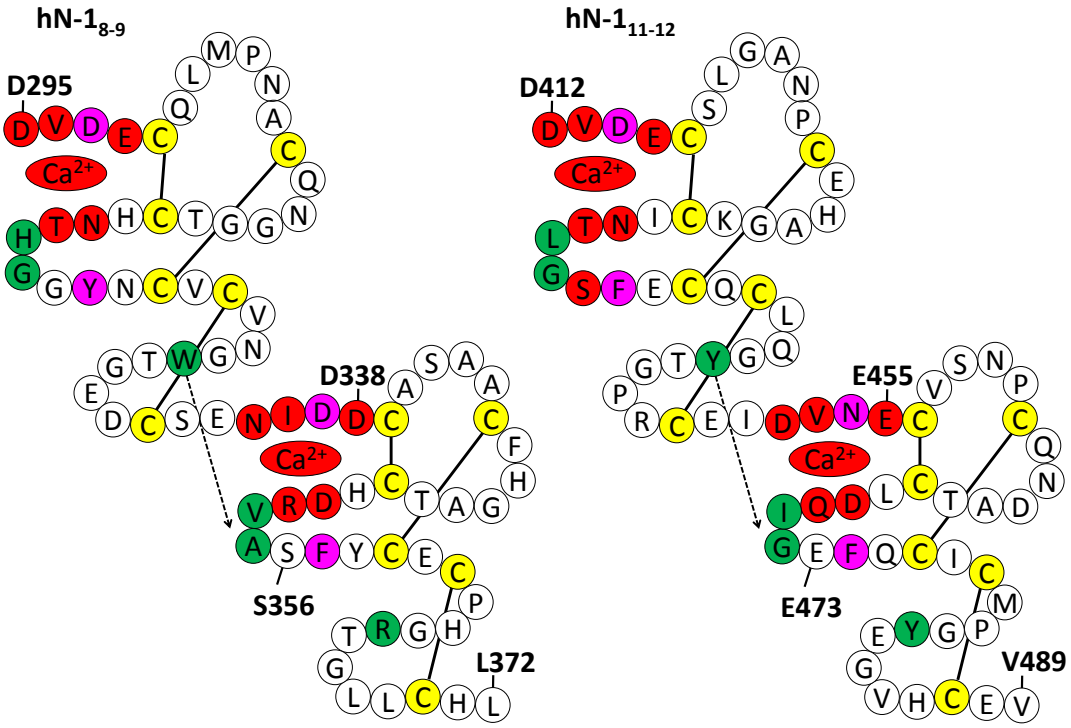


Figure 8.26 Schematic diagrams of EGF8-9 (left hand panel) and EGF11-12 (right hand panel) are shown. Both constructs have the same number of residues in the C1-C2 loops and in the beta-sheet between C3 and C4. An evident difference is the Ca^{2+} binding ligand before C1 in EGF9 (D338) which is an aspartic acid as compared to EGF12 where it is a glutamic acid (E455), the structural effect of which was demonstrated in section 6.5.6. In the schematic diagram, calcium and ligands are depicted in red, interdomain packing residues in green, cysteines in yellow, and conserved residues for intradomain stability in purple.

The RDC values for the EGF8-9 pair were fitted to the EGF11-12 structure using a single set of angles, to test the suitability of the domain orientation observed for EGF11-12, with two sets of angles, to determine an optimal orientation, and with the RDC values for EGF8 scaled up by ~1.5, the ratio of the D_a values for EGF9/EGF8, an approach that was used for hN-1_{11-13}12DG (Table 8.13). The error and q value were significantly lower when fitting with two different sets of angles indicating that EGF8-9 have a different interdomain orientation than EGF11-12. The resulting twist and tilt angles for EGF8-9 are 143 ± 7° and 14 ± 3° with errors calculated as described before. Re-scaling the RDC values for EGF8 by ~1.5 leads to a slightly improved fit for the RDC values for both domains but does not have a significant effect on the calculated twist and tilt angles. The scaling of ~1.5 would correspond to a relative order parameter of ~0.7 for

EGF8 which translates to a cone half angle of ~38°. Interestingly, the twist angle of the EGF8-EGF9 interdomain pair (143 ± 7°) is quite similar to the EGF6-EGF7 interface (151.0°). Both interfaces are characterised by the uncommon Ca^{2+} ligand aspartic acid (D260 in EGF7, D338 in EGF9) at the residue just before the first conserved cysteine that causes the typical twist angle (section 6.5.5, and table 6.1 in section 6.4).

Table 8.13 Fitting of EGF8-9 on the crystal coordinates of EGF11-12.

Dom	Mod	Res	Err	Q	D_a (Hz)	R	Theta (°)	Phi (°)	Psi (°)		
8	11_c	19	4.0	0.24	11.7±0.7	0.41±0.04	21±3	115±6	27±6		
9	12_c	14	3.5	0.12	17.9±0.8	0.50±0.04	31±2	135±3	30±5	**Twist (°)**	**Tilt (°)**
8-9	11-12_c	33	18.0	0.37	13.5±0.5	0.51±0.03	29±2	130±3	24±3	119.3	9.68
8-9	11-12_m	33	9.7	0.27	14.9±0.5	0.48±0.03	20±2	108±6	34±6	143±7	14±3
							32±2	133±3	32±4		
8-9	11-12_{sc}	33	7.3	0.20	18.0±0.5	0.45±0.03	21±2	115±4	28±4	141±6	12±2
							32±2	136±3	29±4		

c: Fitting to the crystal structure
m: Fitting with two sets of angles
sc: Fitting with two sets of angles, but scaling up the RDCs of EGF8 by 1.5

A structural model for EGF10-11 was based on EGF22-23 of the hN-1_{21-23} crystal structure, since EGF10 and EGF22 have the same number of residues in the C1-C2 loop and both pairs have the same number of residues in the β-sheet between C3 and C4 (Figure 8.27). In addition, both have the additional backbone carbonyl Ca^{2+} ligand (S435 in EGF11 and G890 in EGF23), suggesting similar interfaces. The RDC values for EGF11 in the hN-1_{8-11} construct were also fitted using the crystal structure of EGF11 in hN-1_{11-13}. Table 8.14 shows the results of fitting the individual domains EGF10 and EGF11, as well as the domain pair to the crystal structure of EGF22-23. For the fit of EGF10, residues 379 and 389 were removed. For the fit of EGF11, flexible residues 417, 418, 419 and 420 were removed from the fitting. The error analysis was performed as described (section 8.5). A good fit is obtained for the RDCs of EGF11 using the hN-1_{11-13} structure indicating that the overall structure of EGF11 is the same in the hN-1_{8-11} and hN-1_{11-13} constructs.

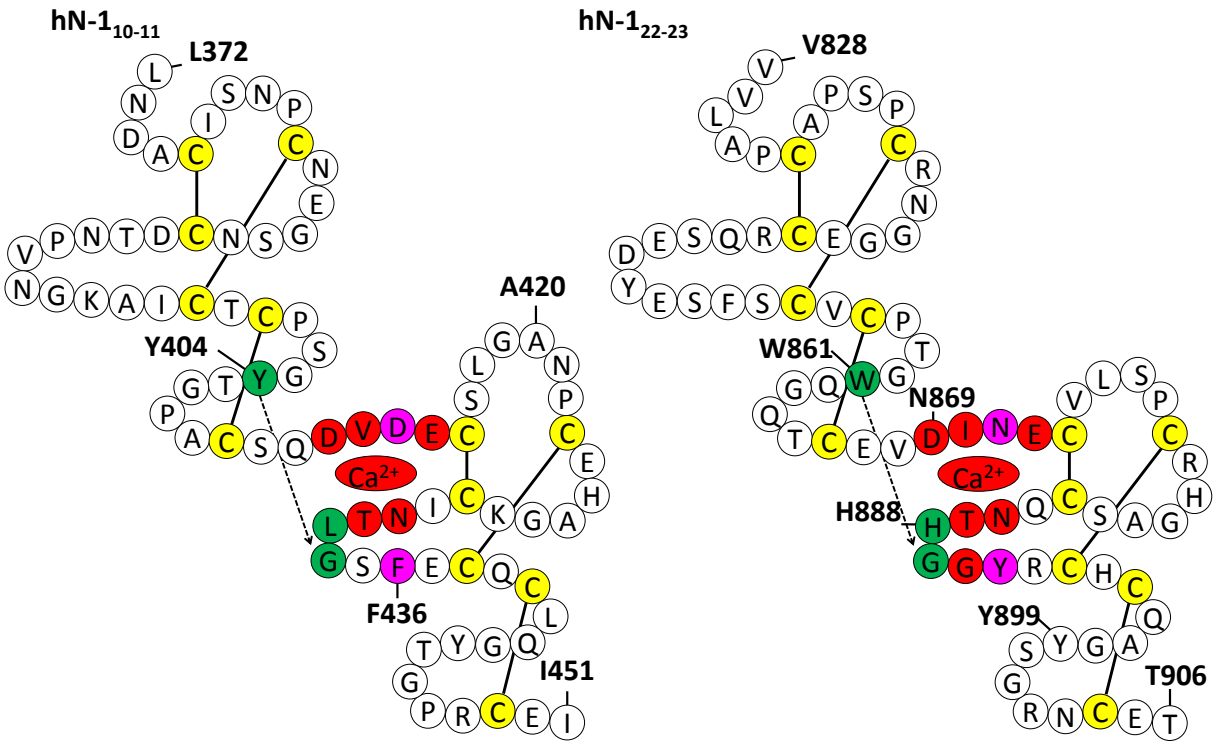


Figure 8.27 Schematic diagrams of EGF10-11 (left hand panel) and EGF22-23 (right hand panel). Note the same number of residues in the C1-C2 and C3-C4 linkers in EGF10 and EGF22; however, in EGF11 there are 6 residues between C1 and C2, but only 4 residues in EGF23. In the schematic diagram, calcium and ligands are depicted in red, interdomain packing residues in green, cysteines in yellow, and conserved residues for intradomain stability in purple.

Table 8.14 Fitting of an alignment tensor to EGF10 and EGF11.

Domain	Model	Res	Err	Q	D_a (Hz)	R	Theta(°)	Phi (°)	Psi (°)		
10	22_{crys}	15	3.2	0.17	-9.2±0.5	0.49±0.13	53±2	67±4	109±6		
11	11_{crys}	26	4.7	0.25	-7.8±0.4	0.44±0.09	97±3	75±2	150±8		
11	23_{crys}	29	6.8	0.30	-7.8±0.4	0.24±0.09	50±2	65±3	95±12	**Twist(°)**	**Tilt(°)**
10-11	22-23	44	6.7	0.28	-8.2±0.3	0.29±0.09	51±2	66±3	103±7	154	55
10-11$_{mod}$	22-23	44	6.3	0.27	-8.3±0.3	0.26±0.08	54±2	67±3	103±7	150±3	55±3
							50±2	64±3	102±16		

crys: Fitting to the crystal structure
mod: Two different sets of angles were used to fit

The D_a value for EGF10 (D_a = -9.2 Hz) was similar to EGF11 (D_a = -7.8 Hz), suggesting rigidity at the interface. Similarly, the angles theta, phi and psi were almost identical suggesting that EGF22-23 is a very good model for the EGF10-11 interface. When allowing two sets of angles, the twist angle only slightly changed from 154° to 150° and the tilt angle did not change at all. This finding is in line with the observation that the additional backbone carbonyl ligand in EGF5

(S201), EGF11 (S435) and EGF23 (G890) causes an increase in twist angle to a value closer to 180°.

The D_a value of 14.9 ± 0.5 Hz (or 18.0 Hz ± 0.5 if scaling is invoked) observed for the EGF8-9 pair is significantly different to the D_a value of -8.3 ± 0.2 Hz observed for the EGF10-11 pair; in particular both the magnitude and sign are different. This suggests that the two pairs of domains align independently in solution; such a result has previously been used to indicate medium- to large-scale interdomain motion (Braddock et al., 2001). This suggests that the EGF9-EGF10 interface is flexible, which is in line with the prediction of EGF10 as a non-calcium binding domain (Hambleton et al., 2004), the absence of an aromatic consensus residue at position four after C5 and the prediction by TALOS+ of lower order parameters for the residues at the beginning of EGF10.

8.7 Discussion

RDC analysis has been performed on the EGF4-13 region of the human Notch-1 receptor. The methodology was first tested using the well-studied hN-$1_{11\text{-}13}$ construct. Good agreement was obtained between the experimental EGF11-13 RDC values and RDC values calculated using the crystal structure, although a slightly wider twist angle was observed in solution. The error analysis of the RDC measurements indicates a limitation in the precision with which twist and tilt angles can be measured. In the hN-$1_{4\text{-}7}$ construct, which was crystallised during this thesis, little variation was found for the EGF4-EGF5 and EGF6-EGF7 interfaces (Table 6.2) suggesting that the error resulting from RDC measurements is not significantly larger than the errors obtained when analysing different crystal structures of the same construct.

RDC analysis on the hN-$1_{11\text{-}13}$12DG construct showed that the abrogation of Ca^{2+} binding in EGF12 has little effect on the structure of EGF11 or EGF13 as the RDCs agreed well with the wildtype structure. The agreement was not good for the RDCs of EGF12 suggesting a different

organisation of this domain. A decrease in D_a was observed for EGF11 when compared to EGF13 suggesting that the two domains do not tumble as a single rigid unit. The Lipari-Szabo formalism was used to explain this finding with a cone model in which EGF11 is able to move freely in a cone of ~45° half-angle with respect to EGF13. This confirms the rigidifying effect of Ca^{2+} on the interdomain organisation observed previously (Downing et al., 1996, Werner et al., 2000, Cordle et al., 2008a, Cordle et al., 2008b). Flexibility was also observed in heteronuclear NOE measurements, with slightly reduced interface NOE values. However, the RDC analysis shows whole domain rearrangements and suggests additional mobility on a slower timescale. Further ^{15}N NMR relaxation or relaxation dispersion experiments will be required to confirm mobility on this slower timescale. Measurements of RDCs for the hN-1_{4-7} construct confirmed the angles observed for the interfaces EGF4-5 and EGF6-7 and only slight differences were found in the solution twist and tilt angles as compared to angles observed in crystallography. However, the tilt angle of the EGF5-EGF6 interface was only ~65° in solution after using RDCs of EGF4-5 and EGF6-7, when compared to 90° observed in the crystal structure. This suggests that this interface may not be entirely rigid and explains the different tilt angles observed in the different crystal structures (section 6.5.1).

RDC analysis of the EGF7-EGF8 interface suggested a twist angle of around 180° which is in line with the twist angles observed for EGF-cbEGF constructs with the extra carbonyl backbone Ca^{2+} ligand. However, the analysis of chemical shifts did not allow the classification of this residue in either the group with the ligand, or the group without (Figure 7.23, section 7.5). Aside from EGF23 (G890), this is possibly the second case where a glycine is able to donate a carbonyl backbone Ca^{2+} ligand at the corresponding position in the sequence. This suggests that the hydroxyl group of serine is not required for this arrangement, but suggests that a small side chain and the ability to adopt an unusual backbone conformation is required (Lee et al., 2004).

The analysis of the EGF8-EGF9 interface showed a smaller D_a for EGF8, when compared to EGF9, but it was possible to fit a common alignment tensor demonstrating a twist angle of 143° and a tilt angle of 14°. The twist angle obtained is similar to the angle measured for the EGF6-EGF7 interface, ~150°, and since both EGF7 and EGF9 have an aspartate at the residue before C1, this suggests twist angles of around ~140-150° for these types of domains. This twist angle is furthermore different from the angle observed for interfaces in which a serine or a glycine at position three before C4 donates a backbone Ca^{2+} ligand. Since this position is adopted by a serine in EGF9 (S356), this suggests an exception of the finding, so far, that all domains with a glycine or serine at position three before C4 donate a backbone carbonyl Ca^{2+} ligand. This is in line with the analysis of chemical shifts (Figure 7.23, section 7.5) that indicated that this residue does not donate a backbone carbonyl Ca^{2+} ligand.

Fitting of the EGF10-EGF11 interface resulted in a similar domain arrangement as seen in EGF22-23 with a twist angle of 150° and a tilt angle of 55°. The D_a obtained for EGF10-11, however, is of opposite sign to the D_a for EGF8-9, suggesting that the two domain pairs do not tumble as a single rigid unit in solution. This suggests flexibility at the EGF9-EGF10 interface.

9 Discussion

9.1 Aim of thesis

Little is known about the cell surface organisation of human Notch-1 and how ligand binding leads to receptor activation. The extracellular region of Notch consists of 36 EGF domains many of which are Ca^{2+} binding. Ca^{2+} has a rigidifying effect on interdomain contacts (Downing et al., 1996), resulting in a rod-like elongated shape for cbEGF domain pairs. Non-calcium binding EGF domains EGF6, EGF10 and EGF22 were predicted to confer flexibility to the receptor (Hambleton et al., 2004) and a jack-knife organisation of Notch-1 was proposed (Figure 9.1) (Xu et al., 2005, Hambleton et al., 2004).

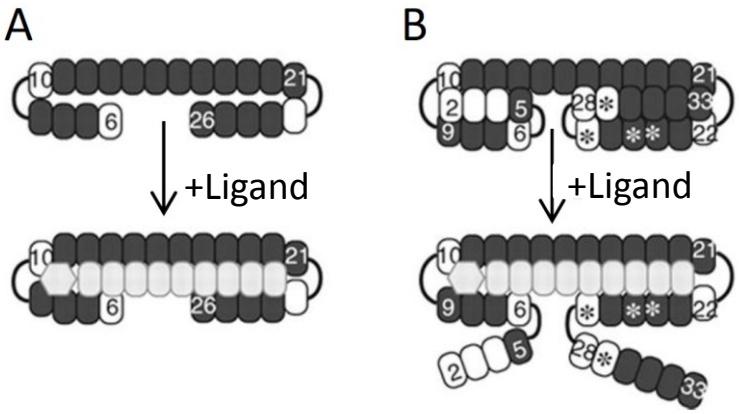


Figure 9.1 Schematic structural models of the Notch-ligand complex by Xu (2005). Domains belonging to the ligand are depicted in light grey, the DSL domain as a hexagon. Domains belonging to Notch are shown in dark grey or in white for constructs predicted to confer flexibility to the receptor. In (A), the ligand is sandwiched in a jack-knife conformation. EGF domains 1-5 and 27-36 were excluded, since they did not increase ligand-binding in this study. In (B), the fact that EGF1-5 slightly decreased ligand binding was used to build a model in which EGF2-5 and EGF28-33 need to be displaced before binding can occur. Note that the MNNL domain at the N-terminus of the DSL domain was not included in this picture. Image was adapted (Xu et al., 2005).

Although it is established that Ca^{2+} binding rigidifies cbEGF-cbEGF domain contacts and that the absence of Ca^{2+} increases their flexibility, it remains unclear whether EGF domains can form well defined interfaces that are different to the ones observed for cbEGF domains. One such example was shown in the crystal structure of factor IX, with a salt bridge between a glutamic acid, one residue further than the aromatic packing residue in the calcium binding EGF1, and an arginine

within the loop between C1-C2 of non-calcium binding EGF2 (Brandstetter et al., 1995), resulting in a 90° angle (section 1.4.5).

This thesis addressed the question, whether EGF6, EGF10 and EGF22 form well defined interfaces or whether they represent sites of flexibility. Furthermore, more detailed analysis of the calcium-binding properties was performed, to identify whether or not EGF domains would adopt a near linear rod-shaped organisation.

9.2 Results

Triple and quadruple EGF domain constructs of the human Notch-1 receptor were expressed in a prokaryotic system, refolded in an *in-vitro* oxido-shuffling system and purified using a nickel-column and chromatography. Ca^{2+} measurements with chromophoric chelation and NMR spectroscopy indicated that under physiological conditions, the Ca^{2+} binding sites studied in this thesis will be saturated (chapter 4). All measurements showed an increase of around two orders of magnitude when an N-terminal EGF domain was added (Figure 4.20) which is in line with previous reports (Smallridge et al., 1999), and confirmed the inter- and intra-domain rigidifying effect of Ca^{2+} (Werner et al., 2000, Downing et al., 1996) – this suggests rod-like interfaces of these regions on the cell-surface as suggested in the model (Figure 9.1).

Crystal structures were obtained in collaboration with Professor Susan Lea at the Sir William Dunn School of Pathology, for the constructs hN-1_{4-7}, hN-1_{21-23} and hN-1_{20-23}. In hN-1_{4-7}, a 90° bend was observed for the EGF5-EGF6 interface, whereas all other interfaces were rod-shaped. The interface between EGF5-EGF6 was defined by interdomain contacts between A208 and V239, and NMR NOESY signals between those residues confirmed that the observed bend is not an artefact of crystal packing. RDC measurements indicated a slightly wider angle in solution (~65-70° instead of 80-90°).

The interface formed by EGF21-22 was rigid and produced a rod-shaped organisation, but a twist angle of 40-50° was found as compared to twist angles of 120-180° observed for the remaining domains. The interface was well defined with an unusual packing between a tyrosine at the turn of the major beta-sheet in EGF22 and a *cis*-proline one residue before the conserved aromatic consensus residue in EGF21.

Attempts to crystallise constructs containing EGF10 were unsuccessful. Heteronuclear NOE experiments gave no hint of fast timescale motions. However, analysis of the random coil shift index predicted small S^2 values for residues between EGF9 and EGF10. In addition, D_a values, observed in RDC measurements, were very different for the EGF8-9 (+14.9±0.9 Hz) and EGF10-11 (-8.3±0.2 Hz) domain pairs. These data suggest flexibility at the EGF9-EGF10 interface. Figure 9.2

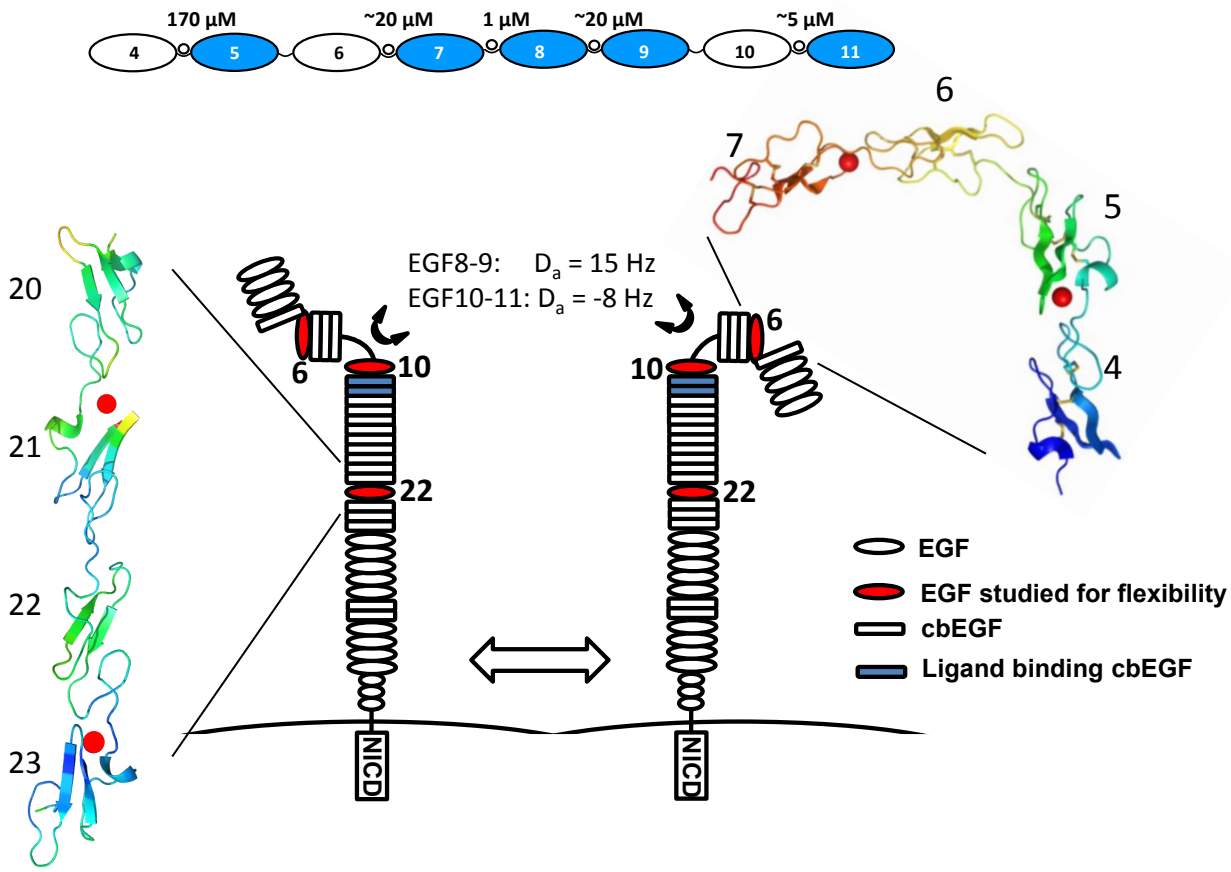


Figure 9.2 Effect of EGF6, EGF10 and EGF22 on the human Notch-1 receptor. The interface between EGF21 and EGF22 is rigid and elongated. A bend was observed between EGF5 and EGF6. Current NMR analysis suggests a flexible interface between EGF9 and EGF10, which is exemplified by the different D_a values obtained. Ca^{2+} is depicted in red.

describes a schematic model of the Notch-1 receptor based on the studies of EGF6, EGF10 and EGF22 and Ca^{2+} measurements, described in this thesis.

Indicated on the top are measured Ca^{2+} affinities. The crystal structures of hN-1_{4-7} and hN-1_{20-23} are shown. Two possible conformations of Notch-1 are displayed based on the flexible interface identified in between EGF9-10. Further characterisation of the interface between EGF9-EGF10 is currently being performed, using T_1/T_2 relaxation and relaxation dispersion experiments. Predictions of the S^2 order parameters in the hN-1_{8-11} construct showed lower S^2 values on the interface between EGF9-EGF10, but also within the major beta-sheet of EGF10. These data also suggest that the EGF9-10 interface is flexible.

Often, flexible regions are of biological importance and mediate protein interactions or enable conformational exchanges. It is possible that the EGF9-10 region might form a tight interface during the interaction with the ligand.

Homology models were generated for the interfaces EGF7-8, EGF8-9 and EGF10-11. The addition of EGF10 to the construct hN-1_{11-13} has been shown to cause a reduction of binding to human Delta-like-1 (Cordle et al., 2008a), an effect that was dependent on Ca^{2+} binding in EGF11. The effect of the absence of Ca^{2+} to an interface formed by a cbEGF domain was described in chapter 8 which resulted in scaling down of the RDC values of EGF11 when Ca^{2+} binding was abrogated in EGF12 of the hN-1_{11-13}12DG mutant. Current research indicates the involvement of L468 and I477 in ligand binding (unpublished). These residues are highlighted on a homology model of hN-1_{10-13} based on the hN-1_{11-13} crystal structure and RDCs collected for hN-1_{8-11} (Figure 9.3). It shows that the kink between EGF10 and EGF11 (tilt angle of ~50°) does not occlude the ligand binding site, assuming a parallel orientation of Notch and ligand during ligand-binding. However, the interaction complex of Notch and ligand is not known in detail and it is possible that both Notch and ligand are coiled around each other rather than forming a parallel

interaction complex. This could explain the function of the 50° tilt angles observed for EGF-cbEGF interfaces with the additional carbonyl backbone Ca^{2+} ligand.

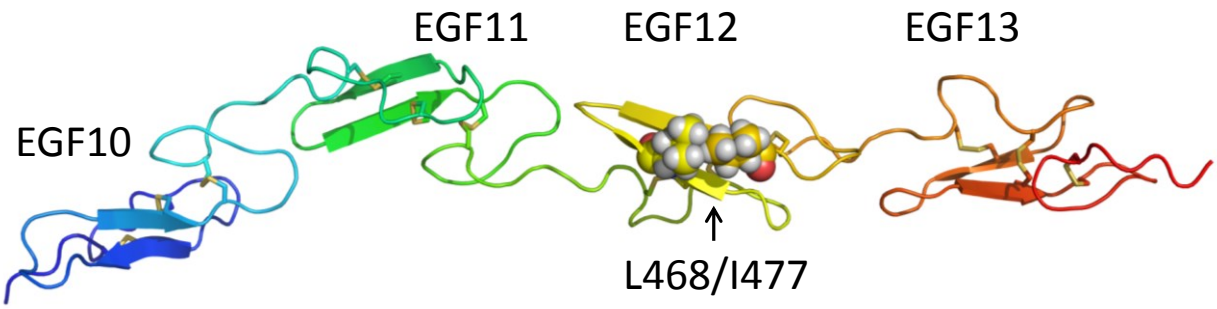


Figure 9.3 shows a model for hN-1_{10-13} based on the crystal structure of hN-1_{11-13}. EGF10 was modelled using the coordinates of EGF22 in the hN-1_{21-23} crystal structure. The interface was defined using RDCs collected for hN-1_{8-11}. Residues involved in ligand binding (L468/I477) are indicated (unpublished). EGF10 points away from the putative ligand binding site.

It is furthermore not known whether EGF10 reduces the binding to Jagged-1/2 which belong to a ligand family with more EGF domains in their extracellular domain when compared to Delta-ligands. The different interdomain tilt and twist angles revealed here for different EGF-EGF pairwise interactions, show four different possibilities of interfaces formed, based on the Ca^{2+} ligands donated. Tilt and twist angle measurements indicated that the presence of an additional carbonyl backbone Ca^{2+} ligand leads to an increase in twist angle to around 180°; with tilt angles of 40-50° measured for EGF22-23, EGF4-5 and EGF10-11.

9.3 Outlook

It should be noted that the constructs studied in this thesis lacked post-translational modifications such as O-glycosylation that are known to regulate Notch activity. Therefore it is possible that these additions may alter the intramolecular organisation of the EGF domain, although unmodified fragments from the EGF11-13 region retain the ability to bind to ligands. The addition of fucose occurs on the major beta-sheet at a serine or threonine, one residue before C3. This residue points away from the interdomain packing regions and is therefore not

likely to affect the interdomain angles described here. However, the addition of GlcNAc can occur on a Thr residue after the conserved aromatic packing residue in between C5 and C6 and could therefore have an effect on interdomain packing. The addition of O-glucose to the loop between C1 and C2 also occurs, but this loop points away from the packing region and would not affect the interdomain orientation.

The structural data presented in this thesis can be tested in functional experiments. Some attempts to do this were made by examining cell aggregation properties of full length *Drosophila* Notch mutants where the EGF5-6 interface was disrupted. These data did not reveal any disruption in the ability of cells expressing a *Drosophila* Notch mutant to interact with Serrate, unlike the L504A mutant. However these experiments only test binding in a semi-quantitative way and do not examine other aspects of Notch biology such as signalling or receptor endocytosis. Mutants could therefore be tested for their ability to disrupt signalling or any other aspect of the Notch biosynthetic pathway. In addition, the consequences of each introduced mutation could be investigated in a limited domain construct using the NMR assignments, to confirm the planned disruption of the interface.

The proposed flexible interface at EGF9-EGF10, could be substituted with a more rigid interface (either EGF21-22, or a pair of calcium binding EGF domains) and the effects tested in functional assays using either limited domain fragments or the full length proteins. Cordle et al. (2008) used a combination of flow cytometry and SPR to identify a negative regulatory effect of EGF10 on DLL-1 binding. Similar experiments could be performed using engineered constructs.

Analysis of the arrangement of calcium ligands in the crystal structures of Notch constructs identified a role for a backbone carbonyl ligand from a serine or glycine three residues before C4 in altering the position of the bound calcium ion and local structure, compared to those domains

which do not have this ligand. It would be interesting to mutate this residue, observe the effect on calcium binding affinity, and activity of the receptor.

9.4 Summary

In summary, biophysical experiments presented in this thesis have established the structural properties of the EGF domains which comprise the major portion of the Notch extracellular domain. These studies highlight the multiple ways in which this domain type can exist within a modular protein. These data facilitate the design of future experiments probing the molecular basis of Notch ligand recognition and events leading to activation.

10 References

ABBOTT, R. J., SPENDLOVE, I., ROVERSI, P., FITZGIBBON, H., KNOTT, V., TERIETE, P., MCDONNELL, J. M., HANDFORD, P. A. & LEA, S. M. 2007. Structural and functional characterization of a novel T cell receptor co-regulatory protein complex, CD97-CD55. *J Biol Chem,* 282**,** 22023-32.

ACAR, M., JAFAR-NEJAD, H., TAKEUCHI, H., RAJAN, A., IBRANI, D., RANA, N. A., PAN, H., HALTIWANGER, R. S. & BELLEN, H. J. 2008. Rumi is a CAP10 domain glycosyltransferase that modifies Notch and is required for Notch signaling. *Cell,* 132**,** 247-58.

AGRAWAL, N., FREDERICK, M. J., PICKERING, C. R., BETTEGOWDA, C., CHANG, K., LI, R. J., FAKHRY, C., XIE, T. X., ZHANG, J., WANG, J., ZHANG, N., EL-NAGGAR, A. K., JASSER, S. A., WEINSTEIN, J. N., TREVIÑO, L., DRUMMOND, J. A., MUZNY, D. M., WU, Y., WOOD, L. D., HRUBAN, R. H., WESTRA, W. H., KOCH, W. M., CALIFANO, J. A., GIBBS, R. A., SIDRANSKY, D., VOGELSTEIN, B., VELCULESCU, V. E., PAPADOPOULOS, N., WHEELER, D. A., KINZLER, K. W. & MYERS, J. N. 2011. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science,* 333**,** 1154-7.

AHIMOU, F., MOK, L. P., BARDOT, B. & WESLEY, C. 2004. The adhesion force of Notch with Delta and the rate of Notch signaling. *J Cell Biol,* 167**,** 1217-29.

ALAGILLE, D., ODIÈVRE, M., GAUTIER, M. & DOMMERGUES, J. P. 1975. Hepatic ductular hypoplasia associated with characteristic facies, vertebral malformations, retarded physical, mental, and sexual development, and cardiac murmur. *J Pediatr,* 86**,** 63-71.

ANDERSON, C. M., ZUCKER, F. H. & STEITZ, T. A. 1979. Space-filling models of kinase clefts and conformation changes. *Science,* 204**,** 375-80.

ARNETT, K. L., HASS, M., MCARTHUR, D. G., ILAGAN, M. X., ASTER, J. C., KOPAN, R. & BLACKLOW, S. C. 2010. Structural and mechanistic insights into cooperative assembly of dimeric Notch transcription complexes. *Nat Struct Mol Biol,* 17**,** 1312-7.

ASTE-AMÉZAGA, M., ZHANG, N., LINEBERGER, J. E., ARNOLD, B. A., TONER, T. J., GU, M., HUANG, L., VITELLI, S., VO, K. T., HAYTKO, P., ZHAO, J. Z., BALEYDIER, F., L'HEUREUX, S., WANG, H., GORDON, W. R., THORYK, E., ANDRAWES, M. B., TIYANONT, K., STEGMAIER, K., ROTI, G., ROSS, K. N., FRANLIN, L. L., WANG, F., CHASTAIN, M., BETT, A. J., AUDOLY, L. P., ASTER, J. C., BLACKLOW, S. C. & HUBER, H. E. 2010. Characterization of Notch1 antibodies that inhibit signaling of both normal and mutated Notch1 receptors. *PLoS One,* 5**,** e9094.

ASTERMARK, J., HOGG, P. J., BJÖRK, I. & STENFLO, J. 1992. Effects of gamma-carboxyglutamic acid and epidermal growth factor-like modules of factor IX on factor X activation. Studies using proteolytic fragments of bovine factor IX. *J Biol Chem,* 267**,** 3249-56.

AUSTIN, J. & KIMBLE, J. 1987. glp-1 is required in the germ line for regulation of the decision between mitosis and meiosis in C. elegans. *Cell,* 51**,** 589-99.

AUSTIN, J. & KIMBLE, J. 1989. Transcript analysis of glp-1 and lin-12, homologous genes required for cell interactions during development of C. elegans. *Cell,* 58**,** 565-71.

BABON, J., MORGAN, W., KELLY, G., ECCLESTON, J., FEENEY, J. & HOLDER, A. 2007. Structural studies on Plasmodium vivax merozoite surface protein-1. *Molecular and Biochemical Parasitology,* 153**,** 31-40.

BARRANTES, I. B., ELIA, A. J., WÜNSCH, K., HRABE DE ANGELIS, M. H., MAK, T. W., ROSSANT, J., CONLON, R. A., GOSSLER, A. & DE LA POMPA, J. L. 1999. Interaction between Notch signalling and Lunatic fringe during somite boundary formation in the mouse. *Curr Biol,* 9**,** 470-80.

BECAM, I., FIUZA, U. M., ARIAS, A. M. & MILÁN, M. 2010. A role of receptor Notch in ligand cis-inhibition in Drosophila. *Curr Biol,* 20**,** 554-60.

BERGGÅRD, T., MIRON, S., ONNERFJORD, P., THULIN, E., AKERFELDT, K. S., ENGHILD, J. J., AKKE, M. & LINSE, S. 2002. Calbindin D28k exhibits properties characteristic of a Ca2+ sensor. *J Biol Chem,* 277**,** 16662-72.

BERJANSKII, M., RILEY, M. & VAN DOREN, S. R. 2002. Hsc70-interacting HPD loop of the J domain of polyomavirus T antigens fluctuates in ps to ns and micros to ms. *J Mol Biol,* 321**,** 503-16.

BERJANSKII, M. & WISHART, D. S. 2006. NMR: prediction of protein flexibility. *Nat Protoc,* 1**,** 683-8.

BERJANSKII, M. V. & WISHART, D. S. 2005. A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc,* 127**,** 14970-1.

BERJANSKII, M. V. & WISHART, D. S. 2008. Application of the random coil index to studying protein flexibility. *J Biomol NMR,* 40**,** 31-48.

BIEKOFSKY, R. R., MARTIN, S. R., BROWNE, J. P., BAYLEY, P. M. & FEENEY, J. 1998. Ca2+ coordination to backbone carbonyl oxygen atoms in calmodulin and other EF-hand proteins: 15N chemical shifts as probes for monitoring individual-site Ca2+ coordination. *Biochemistry,* 37**,** 7617-29.

BOEHR, D. D., MCELHENY, D., DYSON, H. J. & WRIGHT, P. E. 2006. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science,* 313**,** 1638-42.

BORK, P., DOWNING, A. K., KIEFFER, B. & CAMPBELL, I. D. 1996. Structure and distribution of modules in extracellular proteins. *Q Rev Biophys,* 29**,** 119-67.

BRADDOCK, D. T., CAI, M., BABER, J. L., HUANG, Y. & CLORE, G. M. 2001. Rapid identification of medium- to large-scale interdomain motion in modular proteins using dipolar couplings. *J Am Chem Soc,* 123**,** 8634-5.

BRADLEY, P., MISURA, K. M. & BAKER, D. 2005. Toward high-resolution de novo structure prediction for small proteins. *Science,* 309**,** 1868-71.

BRANDSTETTER, H., BAUER, M., HUBER, R., LOLLAR, P. & BODE, W. 1995. X-ray structure of clotting factor IXa: active site and module structure related to Xase activity and hemophilia B. *Proc Natl Acad Sci U S A,* 92**,** 9796-800.

BRAY, S. 1998. Notch signalling in Drosophila: three ways to use a pathway. *Semin Cell Dev Biol,* 9**,** 591-7.

BRÜCKNER, K., PEREZ, L., CLAUSEN, H. & COHEN, S. 2000. Glycosyltransferase activity of Fringe modulates Notch-Delta interactions. *Nature,* 406**,** 411-5.

BRÜNGER, A. T., ADAMS, P. D., CLORE, G. M., DELANO, W. L., GROS, P., GROSSE-KUNSTLEVE, R. W., JIANG, J. S., KUSZEWSKI, J., NILGES, M., PANNU, N. S., READ, R. J., RICE, L. M., SIMONSON, T. & WARREN, G. L. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr,* 54**,** 905-21.

CARDY, C. M. & HANDFORD, P. A. 1998. Metal ion dependency of microfibrils supports a rod-like conformation for fibrillin-1 calcium-binding epidermal growth factor-like domains. *J Mol Biol,* 276**,** 855-60.

CHEN, J., MOLONEY, D. J. & STANLEY, P. 2001. Fringe modulation of Jagged1-induced Notch signaling requires the action of beta 4galactosyltransferase-1. *Proc Natl Acad Sci U S A,* 98**,** 13716-21.

CHILLAKURI, C. R., SHEPPARD, D., LEA, S. M. & HANDFORD, P. A. 2012. Notch receptor-ligand binding and activation: insights from molecular studies. *Semin Cell Dev Biol,* 23**,** 421-8.

CHOU, J. J., LI, S., KLEE, C. B. & BAX, A. 2001. Solution structure of Ca(2+)-calmodulin reveals flexible hand-like properties of its domains. *Nat Struct Biol,* 8**,** 990-7.

CHOU, Y. T., SWAIN, J. F. & GIERASCH, L. M. 2002. Functionally significant mobile regions of Escherichia coli SecA ATPase identified by NMR. *J Biol Chem,* 277**,** 50985-90.

COOKE, R. M., WILKINSON, A. J., BARON, M., PASTORE, A., TAPPIN, M. J., CAMPBELL, I. D., GREGORY, H. & SHEARD, B. 1987. The solution structure of human epidermal growth factor. *Nature,* 327**,** 339-41.

CORDLE, J., JOHNSON, S., TAY, J. Z., ROVERSI, P., WILKIN, M. B., DE MADRID, B. H., SHIMIZU, H., JENSEN, S., WHITEMAN, P., JIN, B., REDFIELD, C., BARON, M., LEA, S. M. & HANDFORD, P. A. 2008a. A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition. *Nat Struct Mol Biol,* 15**,** 849-57.

CORDLE, J., REDFIELD, C., STACEY, M., VAN DER MERWE, P. A., WILLIS, A. C., CHAMPION, B. R., HAMBLETON, S. & HANDFORD, P. A. 2008b. Localization of the delta-like-1-binding site in human Notch-1 and its modulation by calcium affinity. *J Biol Chem,* 283**,** 11785-93.

CORNILESCU, G., MARQUARDT, J., OTTIGER, M. & BAX, A. 1998. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *Journal of the American Chemical Society,* 120**,** 6836-6837.

DAHLBÄCK, B., HILDEBRAND, B. & LINSE, S. 1990. Novel type of very high affinity calcium-binding sites in beta-hydroxyasparagine-containing epidermal growth factor-like domains in vitamin K-dependent protein S. *J Biol Chem,* 265**,** 18481-9.

DANKS, D. M., CAMPBELL, P. E., JACK, I., ROGERS, J. & SMITH, A. L. 1977. Studies of the aetiology of neonatal hepatitis and biliary atresia. *Arch Dis Child,* 52**,** 360-7.

DAVIS, J. A., HANDFORD, P. A. & REDFIELD, C. 2007. The N1317H substitution associated with Leber congenital amaurosis results in impaired interdomain packing in human CRB1 epidermal growth factor-like (EGF) domains. *J Biol Chem,* 282**,** 28807-14.

DE CELIS, J. F. & BRAY, S. 1997. Feed-back mechanisms affecting Notch activation at the dorsoventral boundary in the Drosophila wing. *Development,* 124**,** 3241-51.

DE CELIS, J. F. & BRAY, S. J. 2000. The Abruptex domain of Notch regulates negative interactions between Notch, its ligands and Fringe. *Development,* 127**,** 1291-302.

DE LA POMPA, J. L., WAKEHAM, A., CORREIA, K. M., SAMPER, E., BROWN, S., AGUILERA, R. J., NAKANO, T., HONJO, T., MAK, T. W., ROSSANT, J. & CONLON, R. A. 1997. Conservation of the Notch signalling pathway in mammalian neurogenesis. *Development,* 124**,** 1139-48.

DELAGLIO, F., GRZESIEK, S., VUISTER, G., ZHU, G., PFEIFER, J. & BAX, A. 1995. NMRPIPE - A MULTIDIMENSIONAL SPECTRAL PROCESSING SYSTEM BASED ON UNIX PIPES. *Journal of Biomolecular Nmr,* 6**,** 277-293.

DEXTER, J. S. 1914. The analysis of a case of continuous variation in *Drosophila* by a study of its linkage relations. Am Nat.

DEY, A. & SZOSZKIEWICZ, R. 2012. Complete noise analysis of a simple force spectroscopy AFM setup and its applications to study nanomechanics of mammalian Notch 1 protein. *Nanotechnology,* 23**,** 175101.

DOE, C. & GOODMAN, C. 1985. EARLY EVENTS IN INSECT NEUROGENESIS .2. THE ROLE OF CELL-INTERACTIONS AND CELL LINEAGE IN THE DETERMINATION OF NEURONAL PRECURSOR CELLS. *Developmental Biology,* 111**,** 206-219.

DONOVIEL, D. B., HADJANTONAKIS, A. K., IKEDA, M., ZHENG, H., HYSLOP, P. S. & BERNSTEIN, A. 1999. Mice lacking both presenilin genes exhibit early embryonic patterning defects. *Genes Dev,* 13**,** 2801-10.

DOSSET, P., HUS, J. C., MARION, D. & BLACKLEDGE, M. 2001. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J Biomol NMR,* 20**,** 223-31.

DOWNING, A. K., KNOTT, V., WERNER, J. M., CARDY, C. M., CAMPBELL, I. D. & HANDFORD, P. A. 1996. Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell,* 85**,** 597-605.

ELLISEN, L. W., BIRD, J., WEST, D. C., SORENG, A. L., REYNOLDS, T. C., SMITH, S. D. & SKLAR, J. 1991. TAN-1, the human homolog of the Drosophila notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell,* 66**,** 649-61.

FABBRI, G., RASI, S., ROSSI, D., TRIFONOV, V., KHIABANIAN, H., MA, J., GRUNN, A., FANGAZIO, M., CAPELLO, D., MONTI, S., CRESTA, S., GARGIULO, E., FORCONI, F., GUARINI, A., ARCAINI, L., PAULLI, M., LAURENTI, L., LAROCCA, L. M., MARASCA, R., GATTEI, V., OSCIER, D., BERTONI, F., MULLIGHAN, C. G., FOÁ, R., PASQUALUCCI, L., RABADAN, R., DALLA-FAVERA, R. & GAIDANO, G. 2011. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J Exp Med,* 208**,** 1389-401.

FEHON, R. G., KOOH, P. J., REBAY, I., REGAN, C. L., XU, T., MUSKAVITCH, M. A. & ARTAVANIS-TSAKONAS, S. 1990. Molecular interactions between the protein products of the neurogenic loci Notch and Delta, two EGF-homologous genes in Drosophila. *Cell,* 61**,** 523-34.

FIAUX, J., BERTELSEN, E. B., HORWICH, A. L. & WÜTHRICH, K. 2002. NMR analysis of a 900K GroEL GroES complex. *Nature,* 418**,** 207-11.

FISCHER, M., YEN, W. C., KAPOUN, A. M., WANG, M., O'YOUNG, G., LEWICKI, J., GURNEY, A. & HOEY, T. 2011. Anti-DLL4 inhibits growth and reduces tumor-initiating cell frequency in colorectal tumors with oncogenic KRAS mutations. *Cancer Res,* 71**,** 1520-5.

FLEMING, R. J. 1998. Structural conservation of Notch receptors and ligands. *Seminars in Cell & Developmental Biology,* 9**,** 599-607.

GERSTEIN, M., LESK, A. & CHOTHIA, C. 1994. STRUCTURAL MECHANISMS FOR DOMAIN MOVEMENTS IN PROTEINS. *Biochemistry,* 33**,** 6739-6749.

GLANVILLE, R. W., QIAN, R. Q., MCCLURE, D. W. & MASLEN, C. L. 1994. Calcium binding, hydroxylation, and glycosylation of the precursor epidermal growth factor-like domains of fibrillin-1, the Marfan gene protein. *J Biol Chem,* 269**,** 26630-4.

GLITTENBERG, M., PITSOULI, C., GARVEY, C., DELIDAKIS, C. & BRAY, S. 2006. Role of conserved intracellular motifs in Serrate signalling, cis-inhibition and endocytosis. *EMBO J,* 25**,** 4697-706.

GOLOVANOV, A., HAUTBERGUE, G., WILSON, S. & LIAN, L. 2004. A simple method for improving protein solubility and long-term stability. *Journal of the American Chemical Society,* 126**,** 8933-8939.

GORDON, W. R., ARNETT, K. L. & BLACKLOW, S. C. 2008. The molecular logic of Notch signaling--a structural and biochemical perspective. *J Cell Sci,* 121**,** 3109-19.

GORDON, W. R., ROY, M., VARDAR-ULU, D., GARFINKEL, M., MANSOUR, M. R., ASTER, J. C. & BLACKLOW, S. C. 2009a. Structure of the Notch1-negative regulatory region: implications for normal activation and pathogenic signaling in T-ALL. *Blood,* 113**,** 4381-90.

GORDON, W. R., VARDAR-ULU, D., HISTEN, G., SANCHEZ-IRIZARRY, C., ASTER, J. C. & BLACKLOW, S. C. 2007. Structural basis for autoinhibition of Notch. *Nat Struct Mol Biol,* 14**,** 295-300.

GORDON, W. R., VARDAR-ULU, D., L'HEUREUX, S., ASHWORTH, T., MALECKI, M. J., SANCHEZ-IRIZARRY, C., MCARTHUR, D. G., HISTEN, G., MITCHELL, J. L., ASTER, J. C. & BLACKLOW, S. C. 2009b. Effects of S1 cleavage on the structure, surface export, and signaling activity of human Notch1 and Notch2. *PLoS One,* 4**,** e6613.

GREENWALD, I. 1998. LIN-12/Notch signaling: lessons from worms and flies. *Genes Dev,* 12**,** 1751-62.

GREENWALD, I. & SEYDOUX, G. 1990. Analysis of gain-of-function mutations of the lin-12 gene of Caenorhabditis elegans. *Nature,* 346**,** 197-9.

GREENWALD, I. S., STERNBERG, P. W. & HORVITZ, H. R. 1983. The lin-12 locus specifies cell fates in Caenorhabditis elegans. *Cell,* 34**,** 435-44.

GROTH, C. & FORTINI, M. E. 2012. Therapeutic approaches to modulating Notch signaling: current challenges and future prospects. *Semin Cell Dev Biol,* 23**,** 465-72.

GURUHARSHA, K. G., KANKEL, M. W. & ARTAVANIS-TSAKONAS, S. 2012. The Notch signalling system: recent insights into the complexity of a conserved pathway. *Nat Rev Genet,* 13**,** 654-66.

HAMBLETON, S., VALEYEV, N. V., MURANYI, A., KNOTT, V., WERNER, J. M., MCMICHAEL, A. J., HANDFORD, P. A. & DOWNING, A. K. 2004. Structural and functional properties of the human notch-1 ligand binding region. *Structure,* 12**,** 2173-83.

HANDFORD, P., DOWNING, A. K., RAO, Z., HEWETT, D. R., SYKES, B. C. & KIELTY, C. M. 1995. The calcium binding properties and molecular organization of epidermal growth factor-like domains in human fibrillin-1. *J Biol Chem,* 270**,** 6751-6.

HANDFORD, P. A., BARON, M., MAYHEW, M., WILLIS, A., BEESLEY, T., BROWNLEE, G. G. & CAMPBELL, I. D. 1990. The first EGF-like domain from human factor IX contains a high-affinity calcium binding site. *EMBO J,* 9**,** 475-80.

HANDFORD, P. A., MAYHEW, M., BARON, M., WINSHIP, P. R., CAMPBELL, I. D. & BROWNLEE, G. G. 1991. Key residues involved in calcium-binding motifs in EGF-like domains. *Nature,* 351**,** 164-7.

HERREMAN, A., HARTMANN, D., ANNAERT, W., SAFTIG, P., CRAESSAERTS, K., SERNEELS, L., UMANS, L., SCHRIJVERS, V., CHECLER, F., VANDERSTICHELE, H., BAEKELANDT, V., DRESSEL, R., CUPERS, P., HUYLEBROECK, D., ZWIJSEN, A., VAN LEUVEN, F. & DE STROOPER, B. 1999. Presenilin 2 deficiency causes a mild pulmonary phenotype and no changes in amyloid precursor protein processing but enhances the embryonic lethal phenotype of presenilin 1 deficiency. *Proc Natl Acad Sci U S A,* 96**,** 11872-7.

HIRUMA-SHIMIZU, K., HOSOGUCHI, K., LIU, Y., FUJITANI, N., OHTA, T., HINOU, H., MATSUSHITA, T., SHIMIZU, H., FEIZI, T. & NISHIMURA, S. 2010. Chemical synthesis, folding, and structural insights into O-fucosylated epidermal growth factor-like repeat 12 of mouse Notch-1 receptor. *J Am Chem Soc,* 132**,** 14857-65.

HOEY, T., YEN, W. C., AXELROD, F., BASI, J., DONIGIAN, L., DYLLA, S., FITCH-BRUHNS, M., LAZETIC, S., PARK, I. K., SATO, A., SATYAL, S., WANG, X., CLARKE, M. F., LEWICKI, J. & GURNEY, A. 2009. DLL4 blockade inhibits tumor growth and reduces tumor-initiating cell frequency. *Cell Stem Cell,* 5**,** 168-77.

HUANG, L. H., KE, X. H., SWEENEY, W. & TAM, J. P. 1989. Calcium binding and putative activity of the epidermal growth factor domain of blood coagulation factor IX. *Biochem Biophys Res Commun,* 160**,** 133-9.

INCE, N., DE LA MONTE, S. M. & WANDS, J. R. 2000. Overexpression of human aspartyl (asparaginyl) beta-hydroxylase is associated with malignant transformation. *Cancer Res,* 60**,** 1261-6.

IRVINE, K. D. 2008. A notch sweeter. *Cell,* 132**,** 177-9.

JAENICKE, R. & RUDOLPH, R. 1989. FOLDING PROTEINS. *Creighton, T. E. (Ed.). Practical Approach Series: Protein Structure: a Practical Approach. Xviii+355p. Irl Press at Oxford University Press: New York, New York, USA; Oxford, England, Uk. Illus***,** 191-224.

JENSEN, S. A., CORBETT, A. R., KNOTT, V., REDFIELD, C. & HANDFORD, P. A. 2005. Ca2+-dependent interface formation in fibrillin-1. *J Biol Chem,* 280**,** 14076-84.

KAMATH, B. M., BAUER, R. C., LOOMES, K. M., CHAO, G., GERFEN, J., HUTCHINSON, A., HARDIKAR, W., HIRSCHFIELD, G., JARA, P., KRANTZ, I. D., LAPUNZINA, P., LEONARD, L., LING, S., NG, V. L., HOANG, P. L., PICCOLI, D. A. & SPINNER, N. B. 2012. NOTCH2 mutations in Alagille syndrome. *J Med Genet,* 49**,** 138-44.

KELLY, D. F., LAKE, R. J., MIDDELKOOP, T. C., FAN, H. Y., ARTAVANIS-TSAKONAS, S. & WALZ, T. 2010. Molecular structure and dimeric organization of the Notch extracellular domain as revealed by electron microscopy. *PLoS One,* 5**,** e10532.

KETTLE, S., YUAN, X., GRUNDY, G., KNOTT, V., DOWNING, A. K. & HANDFORD, P. A. 1999. Defective calcium binding to fibrillin-1: consequence of an N2144S change for fibrillin-1 structure and function. *J Mol Biol,* 285**,** 1277-87.

KLINAKIS, A., LOBRY, C., ABDEL-WAHAB, O., OH, P., HAENO, H., BUONAMICI, S., VAN DE WALLE, I., CATHELIN, S., TRIMARCHI, T., ARALDI, E., LIU, C., IBRAHIM, S., BERAN, M., ZAVADIL, J., EFSTRATIADIS, A., TAGHON, T., MICHOR, F., LEVINE, R. L. & AIFANTIS, I. 2011. A novel tumour-suppressor function for the Notch pathway in myeloid leukaemia. *Nature,* 473**,** 230-3.

KNOTT, V., DOWNING, A. K., CARDY, C. M. & HANDFORD, P. 1996. Calcium binding properties of an epidermal growth factor-like domain pair from human fibrillin-1. *J Mol Biol,* 255**,** 22-7.

KOVALL, R. A. & HENDRICKSON, W. A. 2004. Crystal structure of the nuclear effector of Notch signaling, CSL, bound to DNA. *EMBO J,* 23**,** 3441-51.

KRAMER, F., DESHMUKH, M., KESSLER, H. & GLASER, S. 2004. Residual dipolar coupling constants: An elementary derivation of key equations. *Concepts in Magnetic Resonance Part a,* 21A**,** 10-21.

KRIDEL, R., MEISSNER, B., ROGIC, S., BOYLE, M., TELENIUS, A., WOOLCOCK, B., GUNAWARDANA, J., JENKINS, C., COCHRANE, C., BEN-NERIAH, S., TAN, K., MORIN, R. D., OPAT, S., SEHN, L. H., CONNORS, J. M., MARRA, M. A., WENG, A. P., STEIDL, C. & GASCOYNE, R. D. 2012. Whole transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. *Blood,* 119**,** 1963-71.

LAEMMLI, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227**,** 680-5.

LARDELLI, M., DAHLSTRAND, J. & LENDAHL, U. 1994. The novel Notch homologue mouse Notch 3 lacks specific epidermal growth factor-repeats and is expressed in proliferating neuroepithelium. *Mech Dev,* 46**,** 123-36.

LAVAISSIERE, L., JIA, S., NISHIYAMA, M., DE LA MONTE, S., STERN, A. M., WANDS, J. R. & FRIEDMAN, P. A. 1996. Overexpression of human aspartyl(asparaginyl)beta-hydroxylase in hepatocellular carcinoma and cholangiocarcinoma. *J Clin Invest,* 98**,** 1313-23.

LEE, S. S., KNOTT, V., JOVANOVIĆ, J., HARLOS, K., GRIMES, J. M., CHOULIER, L., MARDON, H. J., STUART, D. I. & HANDFORD, P. A. 2004. Structure of the integrin binding fragment from fibrillin-1 gives new insights into microfibril organization. *Structure,* 12**,** 717-29.

LEHMANN, R., JIMENEZ, F., DIETRICH, U. & CAMPOSORTEGA, J. 1983. ON THE PHENOTYPE AND DEVELOPMENT OF MUTANTS OF EARLY NEUROGENESIS IN DROSOPHILA-MELANOGASTER. *Wilhelm Rouxs Archives of Developmental Biology,* 192**,** 62-74.

LEI, L., XU, A., PANIN, V. M. & IRVINE, K. D. 2003. An O-fucose site in the ligand binding domain inhibits Notch activation. *Development,* 130**,** 6411-21.

LI, K., LI, Y., WU, W., GORDON, W. R., CHANG, D. W., LU, M., SCOGGIN, S., FU, T., VIEN, L., HISTEN, G., ZHENG, J., MARTIN-HOLLISTER, R., DUENSING, T., SINGH, S., BLACKLOW, S. C., YAO, Z., ASTER, J. C. & ZHOU, B. B. 2008. Modulation of Notch signaling by antibodies specific for the extracellular negative regulatory region of NOTCH3. *J Biol Chem,* 283**,** 8046-54.

LI, L., KRANTZ, I. D., DENG, Y., GENIN, A., BANTA, A. B., COLLINS, C. C., QI, M., TRASK, B. J., KUO, W. L., COCHRAN, J., COSTA, T., PIERPONT, M. E., RAND, E. B., PICCOLI, D. A., HOOD, L. & SPINNER, N. B. 1997. Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nat Genet,* 16**,** 243-51.

LIEBER, T., KIDD, S., ALCAMO, E., CORBIN, V. & YOUNG, M. W. 1993. Antineurogenic phenotypes induced by truncated Notch proteins indicate a role in signal transduction and may point to a novel function for Notch in nuclei. *Genes Dev,* 7**,** 1949-65.

LINSE, S., HELMERSSON, A. & FORSEN, S. 1991. CALCIUM-BINDING TO CALMODULIN AND ITS GLOBULAR DOMAINS. *Journal of Biological Chemistry,* 266**,** 8050-8054.

LIPSITZ, R. & TJANDRA, N. 2004. Residual dipolar couplings in NMR structure analysis. *Annual Review of Biophysics and Biomolecular Structure,* 33**,** 387-413.

LOUVI, A. & ARTAVANIS-TSAKONAS, S. 2006. Notch signalling in vertebrate neural development. *Nat Rev Neurosci,* 7**,** 93-102.

LUBMAN, O. Y., ILAGAN, M. X., KOPAN, R. & BARRICK, D. 2007. Quantitative dissection of the Notch:CSL interaction: insights into the Notch-mediated transcriptional switch. *J Mol Biol,* 365**,** 577-89.

LUO, Y. & HALTIWANGER, R. S. 2005. O-fucosylation of notch occurs in the endoplasmic reticulum. *J Biol Chem,* 280**,** 11289-94.

MA, S. K., WAN, T. S. & CHAN, L. C. 1999. Cytogenetics and molecular genetics of childhood leukemia. *Hematol Oncol,* 17**,** 91-105.

MALBY, S., PICKERING, R., SAHA, S., SMALLRIDGE, R., LINSE, S. & DOWNING, A. K. 2001. The first epidermal growth factor-like domain of the low-density lipoprotein receptor contains a noncanonical calcium binding site. *Biochemistry,* 40**,** 2555-63.

MALECKI, M. J., SANCHEZ-IRIZARRY, C., MITCHELL, J. L., HISTEN, G., XU, M. L., ASTER, J. C. & BLACKLOW, S. C. 2006. Leukemia-associated mutations within the NOTCH1 heterodimerization domain fall into at least two distinct mechanistic classes. *Mol Cell Biol,* 26**,** 4642-51.

MASKOS, K., FERNANDEZ-CATALAN, C., HUBER, R., BOURENKOV, G. P., BARTUNIK, H., ELLESTAD, G. A., REDDY, P., WOLFSON, M. F., RAUCH, C. T., CASTNER, B. J., DAVIS, R., CLARKE, H. R., PETERSEN, M., FITZNER, J. N., CERRETTI, D. P., MARCH, C. J., PAXTON, R. J., BLACK, R. A. & BODE, W. 1998. Crystal structure of the catalytic domain of human tumor necrosis factor-alpha-converting enzyme. *Proc Natl Acad Sci U S A,* 95**,** 3408-12.

MAYHEW, M., HANDFORD, P., BARON, M., TSE, A. G., CAMPBELL, I. D. & BROWNLEE, G. G. 1992. Ligand requirements for Ca2+ binding to EGF-like domains. *Protein Eng,* 5**,** 489-94.

MCCRIGHT, B., LOZIER, J. & GRIDLEY, T. 2002. A mouse model of Alagille syndrome: Notch2 as a genetic modifier of Jag1 haploinsufficiency. *Development,* 129**,** 1075-82.

MOLONEY, D. J., PANIN, V. M., JOHNSTON, S. H., CHEN, J., SHAO, L., WILSON, R., WANG, Y., STANLEY, P., IRVINE, K. D., HALTIWANGER, R. S. & VOGT, T. F. 2000a. Fringe is a glycosyltransferase that modifies Notch. *Nature,* 406**,** 369-75.

MOLONEY, D. J., SHAIR, L. H., LU, F. M., XIA, J., LOCKE, R., MATTA, K. L. & HALTIWANGER, R. S. 2000b. Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. *J Biol Chem,* 275**,** 9604-11.

MORGAN, T. H. & BRIDGES, C. B. 1916. Sex-linked inheritance in Drosophila. 237**,** 1-88.

MURANYI, A., HAMBLETON, S., KNOTT, V., MCMICHAEL, A., HANDFORD, P. A. & DOWNING, A. K. 2004. 1H, 13C, and 15N resonance assignments of human Notch-1 calcium binding EGF domains 11-13. *J Biomol NMR,* 29**,** 443-4.

NAM, Y., SLIZ, P., SONG, L., ASTER, J. C. & BLACKLOW, S. C. 2006. Structural basis for cooperativity in recruitment of MAML coactivators to Notch transcription complexes. *Cell,* 124**,** 973-83.

NICHOLS, J. T., MIYAMOTO, A., OLSEN, S. L., D'SOUZA, B., YAO, C. & WEINMASTER, G. 2007. DSL ligand endocytosis physically dissociates Notch1 heterodimers before activating proteolysis can occur. *J Cell Biol,* 176**,** 445-58.

NICOLAS, M., WOLFER, A., RAJ, K., KUMMER, J. A., MILL, P., VAN NOORT, M., HUI, C. C., CLEVERS, H., DOTTO, G. P. & RADTKE, F. 2003. Notch1 functions as a tumor suppressor in mouse skin. *Nat Genet,* 33**,** 416-21.

ODA, T., ELKAHLOUN, A. G., PIKE, B. L., OKAJIMA, K., KRANTZ, I. D., GENIN, A., PICCOLI, D. A., MELTZER, P. S., SPINNER, N. B., COLLINS, F. S. & CHANDRASEKHARAPPA, S. C. 1997. Mutations in the human Jagged1 gene are responsible for Alagille syndrome. *Nat Genet,* 16**,** 235-42.

OHLIN, A. K., LINSE, S. & STENFLO, J. 1988. Calcium binding to the epidermal growth factor homology region of bovine protein C. *J Biol Chem,* 263**,** 7411-7.

OKA, C., NAKANO, T., WAKEHAM, A., DE LA POMPA, J. L., MORI, C., SAKAI, T., OKAZAKI, S., KAWAICHI, M., SHIOTA, K., MAK, T. W. & HONJO, T. 1995. Disruption of the mouse RBP-J kappa gene results in early embryonic death. *Development,* 121**,** 3291-301.

OKAMURA, Y. & SAGA, Y. 2008. Notch signaling is required for the maintenance of enteric neural crest progenitors. *Development,* 135**,** 3555-65.

OTTIGER, M., DELAGLIO, F. & BAX, A. 1998. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *Journal of Magnetic Resonance,* 131**,** 373-378.

PANIN, V. M., PAPAYANNOPOULOS, V., WILSON, R. & IRVINE, K. D. 1997. Fringe modulates Notch-ligand interactions. *Nature,* 387**,** 908-12.

PARKS, A. L., KLUEG, K. M., STOUT, J. R. & MUSKAVITCH, M. A. 2000. Ligand endocytosis drives receptor dissociation and activation in the Notch pathway. *Development,* 127**,** 1373-85.

PARKS, A. L., STOUT, J. R., SHEPARD, S. B., KLUEG, K. M., DOS SANTOS, A. A., PARODY, T. R., VASKOVA, M. & MUSKAVITCH, M. A. 2006. Structure-function analysis of delta trafficking, receptor binding and signaling in Drosophila. *Genetics,* 174**,** 1947-61.

PATEL, N. S., LI, J. L., GENERALI, D., POULSOM, R., CRANSTON, D. W. & HARRIS, A. L. 2005. Up-regulation of delta-like 4 ligand in human tumor vasculature and the role of basal expression in endothelial cell function. *Cancer Res,* 65**,** 8690-7.

PEI, Z. & BAKER, N. E. 2008. Competition between Delta and the Abruptex domain of Notch. *BMC Dev Biol,* 8**,** 4.

PERSSON, E., SELANDER, M., LINSE, S., DRAKENBERG, T., OHLIN, A. K. & STENFLO, J. 1989. Calcium binding to the isolated beta-hydroxyaspartic acid-containing epidermal growth factor-like domain of bovine factor X. *J Biol Chem,* 264**,** 16897-904.

POULSON, D. 1940. The effects of certain X-chromosome deficiencies on the embryonic development of Drosophila melanogaster. J. Exp. Zool.

POULSON, D. F. 1937. *The Embryonic development of Drosophila Melanogaster,* Paris, Hermann.

PUENTE, X. S., PINYOL, M., QUESADA, V., CONDE, L., ORDÓÑEZ, G. R., VILLAMOR, N., ESCARAMIS, G., JARES, P., BEÀ, S., GONZÁLEZ-DÍAZ, M., BASSAGANYAS, L., BAUMANN, T., JUAN, M., LÓPEZ-GUERRA, M., COLOMER, D., TUBÍO, J. M., LÓPEZ, C., NAVARRO, A., TORNADOR, C., AYMERICH, M., ROZMAN, M., HERNÁNDEZ, J. M., PUENTE, D. A., FREIJE, J. M., VELASCO, G., GUTIÉRREZ-FERNÁNDEZ, A., COSTA, D., CARRIÓ, A., GUIJARRO, S., ENJUANES, A., HERNÁNDEZ, L., YAGÜE, J., NICOLÁS, P., ROMEO-CASABONA, C. M., HIMMELBAUER, H., CASTILLO, E., DOHM, J. C., DE SANJOSÉ, S., PIRIS, M. A., DE ALAVA, E., SAN MIGUEL, J., ROYO,

R., GELPÍ, J. L., TORRENTS, D., OROZCO, M., PISANO, D. G., VALENCIA, A., GUIGÓ, R., BAYÉS, M., HEATH, S., GUT, M., KLATT, P., MARSHALL, J., RAINE, K., STEBBINGS, L. A., FUTREAL, P. A., STRATTON, M. R., CAMPBELL, P. J., GUT, I., LÓPEZ-GUILLERMO, A., ESTIVILL, X., MONTSERRAT, E., LÓPEZ-OTÍN, C. & CAMPO, E. 2011. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature,* 475**,** 101-5.

RADTKE, F. & RAJ, K. 2003. The role of Notch in tumorigenesis: Oncogene or tumour suppressor? *Nature Reviews Cancer,* 3**,** 756-767.

RANA, N. A., NITA-LAZAR, A., TAKEUCHI, H., KAKUDA, S., LUTHER, K. B. & HALTIWANGER, R. S. 2011. O-glucose trisaccharide is present at high but variable stoichiometry at multiple sites on mouse Notch1. *J Biol Chem,* 286**,** 31623-37.

RAND, M. D., GRIMM, L. M., ARTAVANIS-TSAKONAS, S., PATRIUB, V., BLACKLOW, S. C., SKLAR, J. & ASTER, J. C. 2000. Calcium depletion dissociates and activates heterodimeric notch receptors. *Mol Cell Biol,* 20**,** 1825-35.

RAND, M. D., LINDBLOM, A., CARLSON, J., VILLOUTREIX, B. O. & STENFLO, J. 1997. Calcium binding to tandem repeats of EGF-like modules. Expression and characterization of the EGF-like modules of human Notch-1 implicated in receptor-ligand interactions. *Protein Sci,* 6**,** 2059-71.

RANGANATHAN, P., WEAVER, K. L. & CAPOBIANCO, A. J. 2011. Notch signalling in solid tumours: a little bit of everything but not all the time. *Nat Rev Cancer,* 11**,** 338-51.

RAO, Z., HANDFORD, P., MAYHEW, M., KNOTT, V., BROWNLEE, G. G. & STUART, D. 1995. The structure of a Ca(2+)-binding epidermal growth factor-like domain: its role in protein-protein interactions. *Cell,* 82**,** 131-41.

RAYA, A., KAWAKAMI, Y., RODRÍGUEZ-ESTEBAN, C., IBAÑES, M., RASSKIN-GUTMAN, D., RODRÍGUEZ-LEÓN, J., BÜSCHER, D., FEIJÓ, J. A. & IZPISÚA BELMONTE, J. C. 2004. Notch activity acts as a sensor for extracellular calcium during vertebrate left-right determination. *Nature,* 427**,** 121-8.

REBAY, I., FEHON, R. G. & ARTAVANIS-TSAKONAS, S. 1993. Specific truncations of Drosophila Notch define dominant activated and dominant negative forms of the receptor. *Cell,* 74**,** 319-29.

REBAY, I., FLEMING, R. J., FEHON, R. G., CHERBAS, L., CHERBAS, P. & ARTAVANIS-TSAKONAS, S. 1991. Specific EGF repeats of Notch mediate interactions with Delta and Serrate: implications for Notch as a multifunctional receptor. *Cell,* 67**,** 687-99.

REES, D. J., JONES, I. M., HANDFORD, P. A., WALTER, S. J., ESNOUF, M. P., SMITH, K. J. & BROWNLEE, G. G. 1988. The role of beta-hydroxyaspartate and adjacent carboxylate residues in the first EGF domain of human factor IX. *EMBO J,* 7**,** 2053-61.

REINHARDT, D. P., MECHLING, D. E., BOSWELL, B. A., KEENE, D. R., SAKAI, L. Y. & BÄCHINGER, H. P. 1997. Calcium determines the shape of fibrillin. *J Biol Chem,* 272**,** 7368-73.

REYNOLDS, T. C., SMITH, S. D. & SKLAR, J. 1987. Analysis of DNA surrounding the breakpoints of chromosomal translocations involving the beta T cell receptor gene in human lymphoblastic neoplasms. *Cell,* 50**,** 107-17.

ROEHL, H. & KIMBLE, J. 1993. Control of cell fate in C. elegans by a GLP-1 peptide consisting primarily of ankyrin repeats. *Nature,* 364**,** 632-5.

RUCKERT, M. & OTTING, G. 2000. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *Journal of the American Chemical Society,* 122**,** 7793-7797.

SAHA, S., BOYD, J., WERNER, J. M., KNOTT, V., HANDFORD, P. A., CAMPBELL, I. D. & DOWNING, A. K. 2001. Solution structure of the LDL receptor EGF-AB pair: a paradigm for the assembly of tandem calcium binding EGF domains. *Structure,* 9**,** 451-6.

SANDERS, C. R. & PRESTEGARD, J. H. 1990. Magnetically orientable phospholipid bilayers containing small amounts of a bile salt analogue, CHAPSO. *Biophys J,* 58**,** 447-60.

SANDERS, C. R. & SCHWONEK, J. P. 1992. Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR. *Biochemistry,* 31**,** 8898-905.

SAXENA, A. K., SINGH, K., SU, H. P., KLEIN, M. M., STOWERS, A. W., SAUL, A. J., LONG, C. A. & GARBOCZI, D. N. 2006. The essential mosquito-stage P25 and P28 proteins from Plasmodium form tile-like triangular prisms. *Nat Struct Mol Biol,* 13**,** 90-1.

SCHNEIDER, R., JENSEN, S. A., WHITEMAN, P., MCCULLAGH, J. S., REDFIELD, C. & HANDFORD, P. A. 2010. Biophysical characterisation of fibulin-5 proteins associated with disease. *J Mol Biol,* 401**,** 605-17.

SETHI, M. K., BUETTNER, F. F., KRYLOV, V. B., TAKEUCHI, H., NIFANTIEV, N. E., HALTIWANGER, R. S., GERARDY-SCHAHN, R. & BAKKER, H. 2010. Identification of glycosyltransferase 8 family members as xylosyltransferases acting on O-glucosylated notch epidermal growth factor repeats. *J Biol Chem,* 285**,** 1582-6.

SHEN, Y., DELAGLIO, F., CORNILESCU, G. & BAX, A. 2009. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR,* 44**,** 213-23.

SHEN, Y., LANGE, O., DELAGLIO, F., ROSSI, P., ARAMINI, J. M., LIU, G., ELETSKY, A., WU, Y., SINGARAPU, K. K., LEMAK, A., IGNATCHENKO, A., ARROWSMITH, C. H., SZYPERSKI, T., MONTELIONE, G. T., BAKER, D. & BAX, A. 2008. Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci U S A,* 105**,** 4685-90.

SHI, S., STAHL, M., LU, L. & STANLEY, P. 2005. Canonical Notch signaling is dispensable for early cell fate specifications in mammals. *Mol Cell Biol,* 25**,** 9503-8.

SHI, S. & STANLEY, P. 2003. Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. *Proc Natl Acad Sci U S A,* 100**,** 5234-9.

SHIMIZU, K., CHIBA, S., KUMANO, K., HOSOYA, N., TAKAHASHI, T., KANDA, Y., HAMADA, Y., YAZAKI, Y. & HIRAI, H. 1999. Mouse jagged1 physically interacts with notch2 and other notch receptors. Assessment by quantitative methods. *J Biol Chem,* 274**,** 32961-9.

SHIMIZU, K., CHIBA, S., SAITO, T., KUMANO, K. & HIRAI, H. 2000. Physical interaction of Delta1, Jagged1, and Jagged2 with Notch1 and Notch3 receptors. *Biochem Biophys Res Commun,* 276**,** 385-9.

SMALLRIDGE, R. S., WHITEMAN, P., DOERING, K., HANDFORD, P. A. & DOWNING, A. K. 1999. EGF-like domain calcium affinity modulated by N-terminal domain linkage in human fibrillin-1. *J Mol Biol,* 286**,** 661-8.

SMALLRIDGE, R. S., WHITEMAN, P., WERNER, J. M., CAMPBELL, I. D., HANDFORD, P. A. & DOWNING, A. K. 2003. Solution structure and dynamics of a calcium binding epidermal growth factor-like domain pair from the neonatal region of human fibrillin-1. *J Biol Chem,* 278**,** 12199-206.

SOUTH, A. P., CHO, R. J. & ASTER, J. C. 2012. The double-edged sword of Notch signaling in cancer. *Semin Cell Dev Biol,* 23**,** 458-64.

SPRINZAK, D., LAKHANPAL, A., LEBON, L., SANTAT, L. A., FONTES, M. E., ANDERSON, G. A., GARCIA-OJALVO, J. & ELOWITZ, M. B. 2010. Cis-interactions between Notch and Delta generate mutually exclusive signalling states. *Nature,* 465**,** 86-90.

STENBERG, Y., LINSE, S., DRAKENBERG, T. & STENFLO, J. 1997. The high affinity calcium-binding sites in the epidermal growth factor module region of vitamin K-dependent protein S. *Journal of Biological Chemistry,* 272**,** 23255-23260.

STENFLO, J., OHLIN, A. K., OWEN, W. G. & SCHNEIDER, W. J. 1988. beta-Hydroxyaspartic acid or beta-hydroxyasparagine in bovine low density lipoprotein receptor and in bovine thrombomodulin. *J Biol Chem,* 263**,** 21-4.

STRANSKY, N., EGLOFF, A. M., TWARD, A. D., KOSTIC, A. D., CIBULSKIS, K., SIVACHENKO, A., KRYUKOV, G. V., LAWRENCE, M. S., SOUGNEZ, C., MCKENNA, A., SHEFLER, E., RAMOS, A. H., STOJANOV, P., CARTER, S. L., VOET, D., CORTÉS, M. L., AUCLAIR, D., BERGER, M. F., SAKSENA, G., GUIDUCCI, C., ONOFRIO, R. C., PARKIN, M., ROMKES, M., WEISSFELD, J. L., SEETHALA, R. R., WANG, L., RANGEL-ESCAREÑO, C., FERNANDEZ-LOPEZ, J. C., HIDALGO-MIRANDA, A., MELENDEZ-ZAJGLA, J., WINCKLER, W., ARDLIE, K., GABRIEL, S. B., MEYERSON, M., LANDER, E. S., GETZ, G., GOLUB, T. R., GARRAWAY, L. A. & GRANDIS, J. R. 2011. The mutational landscape of head and neck squamous cell carcinoma. *Science,* 333**,** 1157-60.

STRYNADKA, N. C. & JAMES, M. N. 1989. Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu Rev Biochem,* 58**,** 951-98.

SUK, J. Y., JENSEN, S., MCGETTRICK, A., WILLIS, A. C., WHITEMAN, P., REDFIELD, C. & HANDFORD, P. A. 2004. Structural consequences of cysteine substitutions C1977Y and C1977R in calcium-binding epidermal growth factor-like domain 30 of human fibrillin-1. *J Biol Chem,* 279**,** 51258-65.

SUN, X. & ARTAVANIS-TSAKONAS, S. 1997. Secreted forms of DELTA and SERRATE define antagonists of Notch signaling in Drosophila. *Development,* 124**,** 3439-48.

SUNNERHAGEN, M. S., PERSSON, E., DAHLQVIST, I., DRAKENBERG, T., STENFLO, J., MAYHEW, M., ROBIN, M., HANDFORD, P., TILLEY, J. W. & CAMPBELL, I. D. 1993. The effect of aspartate hydroxylation on calcium binding to epidermal growth factor-like modules in coagulation factors IX and X. *J Biol Chem,* 268**,** 23339-44.

TAKEUCHI, H. & HALTIWANGER, R. S. 2010. Role of glycosylation of Notch in development. *Semin Cell Dev Biol,* 21**,** 638-45.

TAKEUCHI, H., KANTHARIA, J., SETHI, M. K., BAKKER, H. & HALTIWANGER, R. S. 2012. Site-specific O-glucosylation of the epidermal growth factor-like (EGF) repeats of Notch: efficiency of glycosylation is affected by proper folding and amino acid sequence of individual EGF repeats. *J Biol Chem*.

TAMURA, K., TANIGUCHI, Y., MINOGUCHI, S., SAKAI, T., TUN, T., FURUKAWA, T. & HONJO, T. 1995. Physical interaction between a novel domain of the receptor Notch and the transcription factor RBP-J kappa/Su(H). *Curr Biol,* 5**,** 1416-23.

THIM, L., BJOERN, S., CHRISTENSEN, M., NICOLAISEN, E. M., LUND-HANSEN, T., PEDERSEN, A. H. & HEDNER, U. 1988. Amino acid sequence and posttranslational modifications of human factor VIIa from plasma and transfected baby hamster kidney cells. *Biochemistry,* 27**,** 7785-93.

TOLMAN, J. R., FLANAGAN, J. M., KENNEDY, M. A. & PRESTEGARD, J. H. 1995. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc Natl Acad Sci U S A,* 92**,** 9279-83.

TOLMAN, J. R., FLANAGAN, J. M., KENNEDY, M. A. & PRESTEGARD, J. H. 1997. NMR evidence for slow collective motions in cyanometmyoglobin. *Nat Struct Biol,* 4**,** 292-7.

UYTTENDAELE, H., MARAZZI, G., WU, G., YAN, Q., SASSOON, D. & KITAJEWSKI, J. 1996. Notch4/int-3, a mammary proto-oncogene, is an endothelial cell-specific mammalian Notch gene. *Development,* 122**,** 2251-9.

VALCARCE, C., SELANDER-SUNNERHAGEN, M., TÄMLITZ, A. M., DRAKENBERG, T., BJÖRK, I. & STENFLO, J. 1993. Calcium affinity of the NH2-terminal epidermal growth factor-like module of factor X. Effect of the gamma-carboxyglutamic acid-containing module. *J Biol Chem,* 268**,** 26673-8.

VRANKEN, W. F., BOUCHER, W., STEVENS, T. J., FOGH, R. H., PAJON, A., LLINAS, M., ULRICH, E. L., MARKLEY, J. L., IONIDES, J. & LAUE, E. D. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins,* 59**,** 687-96.

WANG, N. J., SANBORN, Z., ARNETT, K. L., BAYSTON, L. J., LIAO, W., PROBY, C. M., LEIGH, I. M., COLLISSON, E. A., GORDON, P. B., JAKKULA, L., PENNYPACKER, S., ZOU, Y., SHARMA, M., NORTH, J. P., VEMULA, S. S., MAURO, T. M., NEUHAUS, I. M., LEBOIT, P. E., HUR, J. S., PARK, K., HUH, N., KWOK, P. Y., ARRON, S. T., MASSION, P. P., BALE, A. E., HAUSSLER, D., CLEAVER, J. E., GRAY, J. W., SPELLMAN, P. T., SOUTH, A. P., ASTER, J. C., BLACKLOW, S. C. & CHO, R. J. 2011. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. *Proc Natl Acad Sci U S A,* 108**,** 17761-6.

WANG, Y. & SPELLMAN, M. W. 1998. Purification and characterization of a GDP-fucose:polypeptide fucosyltransferase from Chinese hamster ovary cells. *J Biol Chem,* 273**,** 8112-8.

WEINMASTER, G., ROBERTS, V. J. & LEMKE, G. 1992. Notch2: a second mammalian Notch gene. *Development,* 116**,** 931-41.

WENG, A. P., FERRANDO, A. A., LEE, W., MORRIS, J. P., SILVERMAN, L. B., SANCHEZ-IRIZARRY, C., BLACKLOW, S. C., LOOK, A. T. & ASTER, J. C. 2004. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science,* 306**,** 269-71.

WERNER, J. M., KNOTT, V., HANDFORD, P. A., CAMPBELL, I. D. & DOWNING, A. K. 2000. Backbone dynamics of a cbEGF domain pair in the presence of calcium. *J Mol Biol,* 296**,** 1065-78.

WHARTON, K. A., JOHANSEN, K. M., XU, T. & ARTAVANIS-TSAKONAS, S. 1985. Nucleotide sequence from the neurogenic locus notch implies a gene product that shares homology with proteins containing EGF-like repeats. *Cell,* 43**,** 567-81.

WHITEMAN, P., DOWNING, A. K. & HANDFORD, P. A. 1998. NMR analysis of cbEGF domains gives new insights into the structural consequences of a P1148A substitution in fibrillin-1. *Protein Eng,* 11**,** 957-9.

WHITEMAN, P., WILLIS, A. C., WARNER, A., BROWN, J., REDFIELD, C. & HANDFORD, P. A. 2007. Cellular and molecular studies of Marfan syndrome mutations identify co-operative protein folding in the cbEGF12-13 region of fibrillin-1. *Hum Mol Genet,* 16**,** 907-18.

WISHART, D. S. & SYKES, B. D. 1994. Chemical shifts as a tool for structure determination. *Methods Enzymol,* 239**,** 363-92.

WRIGHT, T. R. 1970. The genetics of embryogenesis in Drosophila. *Adv Genet,* 15**,** 261-395.

WU, Y., CAIN-HOM, C., CHOY, L., HAGENBEEK, T. J., DE LEON, G. P., CHEN, Y., FINKLE, D., VENOOK, R., WU, X., RIDGWAY, J., SCHAHIN-REED, D., DOW, G. J., SHELTON, A., STAWICKI, S., WATTS, R. J., ZHANG, J., CHOY, R., HOWARD, P., KADYK, L., YAN, M., ZHA, J., CALLAHAN, C. A., HYMOWITZ, S. G. & SIEBEL, C. W. 2010. Therapeutic antibody targeting of individual Notch receptors. *Nature,* 464**,** 1052-7.

XU, A., LEI, L. & IRVINE, K. D. 2005. Regions of Drosophila Notch that contribute to ligand binding and the modulatory influence of Fringe. *J Biol Chem,* 280**,** 30158-65.

XU, T., REBAY, I., FLEMING, R. J., SCOTTGALE, T. N. & ARTAVANIS-TSAKONAS, S. 1990. The Notch locus and the genetic circuitry involved in early Drosophila neurogenesis. *Genes Dev,* 4**,** 464-75.

XUE, Y., GAO, X., LINDSELL, C. E., NORTON, C. R., CHANG, B., HICKS, C., GENDRON-MAGUIRE, M., RAND, E. B., WEINMASTER, G. & GRIDLEY, T. 1999. Embryonic lethality and vascular defects in mice lacking the Notch ligand Jagged1. *Hum Mol Genet,* 8**,** 723-30.

YOCHEM, J. & GREENWALD, I. 1989. glp-1 and lin-12, genes implicated in distinct cell-cell interactions in C. elegans, encode similar transmembrane proteins. *Cell,* 58**,** 553-63.

YOCHEM, J., WESTON, K. & GREENWALD, I. 1988. The Caenorhabditis elegans lin-12 gene encodes a transmembrane protein with overall similarity to Drosophila Notch. *Nature,* 335**,** 547-50.

YUWEN, T., POST, C. & SKRYNNIKOV, N. 2011. Domain cooperativity in multidomain proteins: what can we learn from molecular alignment in anisotropic media? *Journal of Biomolecular Nmr,* 51**,** 131-150.

ZWECKSTETTER, M. 2008. NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc,* 3**,** 679-90.

ZWECKSTETTER, M. & BAX, A. 2000. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *Journal of the American Chemical Society,* 122**,** 3791-3792.

ZWEIFEL, M. E., LEAHY, D. J., HUGHSON, F. M. & BARRICK, D. 2003. Structure and stability of the ankyrin domain of the Drosophila Notch receptor. *Protein Sci,* 12**,** 2622-32.

Appendix

Chemical shifts (ppm) for hN-1_{4-7} (30 mM Ca^{2+}, pH 6.1)

Res	Typ	H	N	ND	NE	HA	HB	HG	HD
140	Gln		120.1			4.32	2.08 1.99	2.39 2.37	
141	Ala	8.36	125.88			4.26	1.34		
142	Asp	8.35	120.9			4.91	2.68 2.83		
143	Pro					4.44	2.43 1.87	2.15 1.97	3.98 3.99
144	Cys	8.43	114.21			4.46	3.23 2.95		
145	Ala	7.59	125.56			4.01	1.48		
146	Ser	7.73	110.39			4.36	3.79 3.79		
147	Asn	8.01	117.53	111.9		4.43	2.90 2.90		6.70 7.42
148	Pro					4.27	1.77 1.64	1.22 0.74	3.47 3.47
149	Cys	7.82	118.11			4.32	2.33 2.78		
150	Ala	8.31	127.4			4.31	1.2		
151	Asn	8.6	111.06			4.03	2.15 2.04		
152	Gly	8.49	103.21			3.64 4.10			
153	Gly	7.88	106.69			3.51 4.32			
154	Gln	8.43	121.84			4.53	2.19 2.09	2.34 2.43	
155	Cys	8.93	127.74			5.04	3.04 3.04		
156	Leu	9.4	132.74			5.11	1.62 1.63	1.65	0.79 0.93
157	Pro					4.66	2.04 2.29		3.91
158	Phe	8.23	123.5			4.5	2.57 2.95		
159	Glu	8.91	126.24			3.49	1.83 1.83	1.65 1.44	
160	Ala	8.21	125.09			4.6	1.44		
161	Ser	8.47	116.27			4.67	3.86 3.86		
162	Tyr	8.38	120.71			5.6	3.17 2.59		
163	Ile	9.27	117.86			4.52	1.67	0.96 0.72 1.33	0.59
164	Cys	8.7	119.09			5.59	2.98 2.58		
165	His	9.56	123.79			5.04	3.12 3.33		
166	Cys	9.22	126.08			5.07	3.27 2.73		
167	Pro						2.2	1.99 1.84	3.22
168	Pro					4.46	2.00 2.40		
169	Ser	7.34	108.04			3.97	3.69 3.69		
170	Phe	8.1	119.08			5.34	3.13 2.33		
171	His	9.79	115.36			5.09	3.06 3.30		
172	Gly	8.07	108.06			3.90 4.82			
173	Pro					4.42	2.47 2.13	2.48	3.74 3.83
174	Thr	8.32	104.36			5.12	5	1.3	
175	Cys	7.56	111.8			4.11	3.70 3.02		
176	Arg	8.51	115.31			4	1.91 1.92	1.26 1.26	3.10 3.10
177	Gln	8.31	118.7			4.63	1.34 1.72	2.27 2.34	
178	Asp	9.06	125.22			4.45	2.45 2.70		
179	Val	7.55	126.96			3.7	1.52	0.52 0.55	
182	Cys					4.2	3.17 3.00		
183	Gly	7.11	104.83			3.62 3.94	0.62		
184	Gln	7.47	118.39		111.89	4.09	1.97 1.83	2.25 2.25	
185	Lys	7.27	117.82			4.65	1.66 1.66	1.20 1.21	1.56 1.56
186	Pro					4.42	1.82 2.28		

Res	Typ	H	N	ND	NE	HA	HB	HG	HD
187	Gly	8.62	110.61			3.76			
188	Leu	7.27	121.35				1.46 1.71		0.12 0.52
189	Cys	8.59	115.58				2.39 2.52		
190	Arg	7.85	121.54			3.89	1.61 1.76		
191	His	8.64	113.75			4.05	2.56		
192	Gly	8.86	107.17			3.77 4.00	0.77		
193	Gly	7.21	106.86			3.40 4.39	0.39		
194	Thr	8.77	116.32			4.45	4	1.07	
195	Cys	8.97	129			5.14	3.04 3.02		
196	His	9.74	129.37			4.86	3.13 3.00		
197	Asn	9.29	125.37			4.83	2.83 2.66		
198	Glu	7.75	122.94			4.81	1.27 1.25	2.01 2 .08	
199	Val	9.42	126.78			3.61	1.96	1.14 1.10	
200	Gly	9.12	120.35			3.03 4.26	0.26		
201	Ser	6.75	110.33			4.45	3.97 3.61		
202	Tyr	8.85	118.34			4.9	3.45 2.93		
203	Arg	9.41	118.88			4.67	1.77 1.79	1.46 1.46	3.09
204	Cys	9.02	119.54			5.62	2.69 2.94		
205	Val	9.35	125.91			4.26	2.13	0.88 0.95	
206	Cys	9.06	127.17			4.81	2.76 3.41		
207	Arg	8.65	120.74			4.42	1.92		
208	Ala	8.57	120.77			4.02	1.41		
209	Thr	7.07	101.63			3.99	4.29	1.22	
210	His	8.25	122.52			5.18	2.70 3.09		
211	Thr	9.34	112.37			4.77	4.16	1.1	
212	Gly	7.95	108.38			4.00 4.85	0.85		
213	Pro					4.01	2.05 2.37		
214	Asn	9.18	114.72			5.53	2.43 3.52		
215	Cys	8.06	113.43						
216	Glu	9.89	129.13			4.29	2.00 2.22		
217	Arg	8.75	123						
218	Pro					4.2	2.11 1.74		
219	Tyr	8.19	123.97			4.25	2.65 2.50		
220	Val	7.96	129.29			4.29	2.04	0.81 0.84	
221	Pro								
222	Cys	8.35	122.95				3.38 3.66		
223	Ser	7.38	115.06				3.70 3.41		
224	Pro								
225	Ser	8.66	114.5						
226	Pro					4.64	1.96 1.96	1.23 1.29	3.69 3.88
227	Cys	7.62	118.28			4.29	2.34 2.34		
228	Gln	8.76	123.97			4.31	2.11 1.64		
229	Asn	9.13	114.33			3.99	1.90 1.10		
230	Gly	8.7	106.61			4.03 .3	0.63		
231	Gly	7.39	105.67			3.41 .4	0.38		
232	Thr	8.48	116.85			4.44	4.09	1.24	
233	Cys	9.08	128.71			5.04	3.26 3.14		
234	Arg	9.39	130.19			5.06	1.85 1.85	1.58 1.66	3.21
235	Pro					4.79	2.16 2.30		
236	Thr	8.58	113.02			4.48	4.6	1.03	
237	Gly	7.74	110.28			4.01 4.22	0.22		

Res	Typ	H	N	ND	NE	HA	HB	HG	HD
238	Asp	8.71	120.8			4.43	2.77 2.68		
239	Val	8.41	107.2			4.73	2.44	0.83 0.84	
240	Thr	7.08	111.15			4.83	4.15	1.11	
241	His	8.28	117.29			5.45	3.16 2.73		
242	Glu	9.03	118.03			4.66	1.96 1.97	2.17 2.17	
243	Cys	9.13	119.42			5.5	3.11 2.68		
244	Ala	9.32	128.67			4.75	1.4		
245	Cys	8.84	122.77			4.82	2.74 3.32		
246	Leu	8.68	124.12			4.61	1.66 1.80	1.64	0.99 1.04
247	Pro					4.35	2.30 1.77		
248	Gly	8.74	109.55			3.56 4.13	0.13		
249	Phe	7.94	118.7			5.72	3.12 2.92		7.02
250	Thr	9.41	115.77			4.83	4.06	0.84	
251	Gly	7.92	109.46			3.86 4.91	0.86		
252	Gln	9.4	123.74		111.72	4.02	2.04 2.17	2.44 2.39	
253	Asn	9.13	114.42	112.88		5.75	3.44 2.54		7.48 6.95
254	Cys	7.9	113.17			4.12	3.50 3.19		
255	Glu	10.3	118.48			3.96	1.97 2.18		
256	Glu	8.72	123.33			4.61	1.83 1.82		
257	Asn	8.82	125.58	112.35		4.57	3.13 2.72		6.85 7.35
258	Ile	7.83	129.04			3.53	1.59	0.97 0.72 1.31	0.72
259	Asp	9.06	131.01			4.27	2.69 2.69		
260	Asp	9.22	127.59			4.74	2.85 2.69		
261	Cys	8.42	118.02			4.14	3.57 2.91		
262	Pro					4.27	2.18 1.69	1.77 2.03	3.47 .3
263	Gly	9.03	113.91			3.74 3.96	0.96		
264	Asn	7.61	116.15	115.94		4.01	2.69 1.93		8.16 6.87
265	Asn	8.18	117.06	111.4		4.85	1.99 2.47		6.73 6.12
266	Cys	8.66	116.21			4.55	2.50 2.69		
267	Lys	8.26	124.65			4.6	1.63 1.52		1.27
268	Asn	8.53	112.11	116.02		3.98	1.78 2.03		6.06 6.81
269	Gly	8.28	102.94			3.56 4.21	0.56		
270	Gly	8.1	107.88			3.49 4.37	0.37		
271	Ala	8.3	126.82			4.61	1.43		
272	Cys	8.94	122.95			4.79	2.79 2.93		
273	Val	9.53	134.27			4.26	2.02	0.99 0.93	
274	Asp	9.11	128.38			4.52	2.51 3.02		
275	Gly	8.31	114.82			3.44 3.80	0.8		
276	Val	9.33	127.59			3.85	2.05	0.99 0.95	
277	Asn	8.8	125.74			4.49	3.16 2.43		7.71 6.51
278	Thr	7.34	108.42			4.41	4.35	0.85	
279	Tyr	7.77	114.8			5.08	2.55 3.27		6.85
280	Asn	8.89	117.08	113.14		5.09	2.55 2.74		6.77 7.53
281	Cys	9.25	118.29			5.53	2.64 3.05		
282	Arg	9.78	127			4.65	1.70 1.95	1.52 1.72	3.26
283	Cys	9.13	127.28			5.13	2.75 3.42		
284	Pro					3.82			
285	Pro					4.32	2.38 1.90	2.10 2.10	3.82 3.82
286	Glu	8.7	114.5			3.95	1.65 1.77	2.02 2.09	
287	Trp	8.05	121.2		127.48	5.08	2.68 3.29		
288	Thr	9.07	113.39			4.91	4.28	1.17	
289	Gly	8.2	107.82			3.83 4.82	0.83		

Res	Typ	H	N	ND	NE	HA	HB	HG	HD
290	Gln	8.59	121.17		110.54	3.79	1.34 1.85	1.38 1.76	
291	Tyr	8.93	113.14			5.39	2.57 3.98		7.12
292	Cys	7.65	110.79			4.4	2.97 3.96		
293	Thr	7.33	104.19			4.18	4.6	1.03	
294	Glu	8.83	128.4			4.57	2.32 2.32	1.77 2.11	

Res	Typ	HE	C	CA CB	CB	CG	CD
140	Gln		175.56	55.78	29.84	33.80	
141	Ala		177	52.38	19.51		
142	Asp			51.44	41.75		
143	Pro		177.41	64.41	33.22	27.58	51.13
144	Cys		176.58	53.90	39.79		
145	Ala		178.32	55.38	18.75		
146	Ser		173.84	57.74	63.11		
147	Asn			53.52	37.72		
148	Pro		176.94	64.08	32.68	26.52	50.71
149	Cys		173.29	53.35	38.80		
150	Ala		177.21	51.40	21.52		
151	Asn		174.55	55.04	37.40		
152	Gly		176.40	45.25			
153	Gly		171.15	46.03			
154	Gln		173.88	54.54	31.23		
155	Cys		174	57.34	38.77		
156	Leu			52.6	44.05	27.20	24.53 25.29
157	Pro		176.46	62.96	32.68		50.23
158	Phe		173.59	58.42	41.88		
159	Glu		175.87	58.44	27.1		
160	Ala		177	52.44	18.94		
161	Ser		173.24	55.88	63.73		
162	Tyr		173.62	56.19	42.88		
163	Ile		174.52	59.84	42.69	17.93 27.25	13.66
164	Cys		174.76	53.48	43.25		
165	His		175.51	53.93	29.77		
166	Cys			52.61	35.99		
167	Pro						
168	Pro		176.54	64.76	31.86		
169	Ser		174.31	58.02	63.70		
170	Phe		173.16	58.01	46.30		
171	His		173.77	54.76	31.58		
172	Gly			44.15			
173	Pro		178.05	65.29	32.16	32.85	49.83
174	Thr		174.32	59.63	69.24	22.67	
175	Cys		174.54	54.40	35.52		
176	Arg		177.23	57.26	31.30	29.26	43.39
177	Gln		175.26	55.23	29.50	34.56	
178	Asp		176.57	55.62	43.53		
179	Val			62.37	33.39	20.08 23.61	
182	Cys		176.99	54.16	35.89		
183	Gly		174.80	45.66			
184	Gln	6.75 7.36	176.14	57.11	30.32		

Res	Typ	HE	C	CA CB	CB	CG	CD
185	Lys	2.87		53.04	33.63	24.53	29.43
186	Pro		178.39	63.55	31.77		
187	Gly		174.59	45.77			
188	Leu		177.49	57.87	42.90		23.61 25.12
189	Cys		173.4	51.98	36.02		
190	Arg		175.61	56.13	31.46		
191	His		176.89	56.75	26.44		
192	Gly		175.53	45.85			
193	Gly		171.56	45.58			
194	Thr		172.33	62.01	71.07		
195	Cys		174	57.80	38.06		
196	His		174.01	55.10	31.55		
197	Asn		174.31	54.26	39.47		
198	Glu		175.66	53.85	34.16	34.58	
199	Val		176.96	65.19	30.30	21.53 21.77	
200	Gly		173.50	47.21			
201	Ser		173.22	56.78	63.33		
202	Tyr		171.76	57.44	33.18		
203	Arg		174.61	54.84	34.10	27.52	
204	Cys		175.11	52.99	40.95		
205	Val		176.48	62.36	32.05	21.00 21.43	
206	Cys		176.89	53.99	38.89		0
207	Arg		176.92	56.93	31.54		
208	Ala		177.04	54.58	19.28		
209	Thr		175.4	62.34	69.28	21.89	
210	His		172.65	57.55	35.61		
211	Thr		173.18	60.12	70.75	19.47	
212	Gly		174.30	43.88			
213	Pro		177.35	65.19	31.92		
214	Asn		175.13	53.07	39.69		
215	Cys		176.46	54.64	36.65		
216	Glu		177.01	58.13	30.42		
217	Arg		173.37	53.35			
218	Pro		176.22	62.84	31.89		
219	Tyr		174.43	58.55	39.12		
220	Val		173.63	58.94	33.67	20.36 20.95	
221	Pro			29.53			
222	Cys		173.83	53.47	37.00		
223	Ser		171.72	54.47	63.81		
224	Pro		176.42				
225	Ser						
226	Pro		176.36	62.54	32.05	25.86	50.63
227	Cys		174	54.17	37.76		
228	Gln		176.29	54.70	31/12/14		
229	Asn		176.52	54.89	36.92		
230	Gly		175.28	45.55			
231	Gly		172.12	45.70			
232	Thr		172.83	62.22	71.25	21.68	
233	Cys		173.89	57.87	38.00		
234	Arg		173.48	53.52	32.86	27.13	43.67
235	Pro		177.58	62.83	32.55		

Res	Typ	HE	C	CA CB	CB	CG	CD
236	Thr		173.44	60.21	69.24	22.64	
237	Gly		172.68	44.72			
238	Asp		177.95	57.42	40.76		
239	Val		175.88	60.70	33.30	19.07 21.80	
240	Thr		173.34	60.49	71.65	21.82	
241	His		172.88	55.43	35.48		
242	Glu		175.25	54.74	33.79	35.70	
243	Cys		173.38	53.91	43.82		
244	Ala		177.19	50.85	18.64		
245	Cys		176.01	53.19	37.41		
246	Leu		174.22	53.88	40.20	28.69	22.47 25.97
247	Pro		176.96	64.21	32.09		
248	Gly		174.10	44.29			
249	Phe	7.2	174.5	56.74	42.56		
250	Thr		172.83	59.81	70.43	21.83	
251	Gly		176.78	43.68			
252	Gln	6.83 7.39	176.67	59.47	29.1	33.82	
253	Asn		175.66	52.33	39.57		
254	Cys		174.65	55.14	36.75		
255	Glu		176.25	58.06	29.80		
256	Glu		174.35	55.24	31.05		
257	Asn		175.66	51.79	38.19		
258	Ile		175.97	62.86	38.86	16.63 29.14	13.93
259	Asp		175.47	55.46	40.51		
260	Asp		176.19	55.64	42.69		
261	Cys		172.92	58.80	41.43		
262	Pro		178.41	66.19	30.22	28.56	50.71
263	Gly		175.21	45.82			
264	Asn		174.14	55.47	39.68		
265	Asn		174.53	52.30	39.54		
266	Cys		175.37	55.21	37.09		
267	Lys	2.82	176.92	53.96	34.88		23.93
268	Asn		174.50	54.91	35.89		
269	Gly		176.30	44.95			
270	Gly		171.31	46.51			
271	Ala		176.15	50.46	21.00		
272	Cys		173.86	57.95			
273	Val		175.77	61.71	33.28	21.11 20.61	
274	Asp		175.31	57.64	42.73		
275	Gly		174.02	44.15			
276	Val		173.7	63.55	31.17	22.04 21.46	
277	Asn		173.31	54.10	38.29		
278	Thr		170.62	60.10	67.28	17.87	
279	Tyr	7.12	172.79	56.30	42.06		
280	Asn		173.63	51.83	42.87		
281	Cys		173.92	53.23	41.69		
282	Arg		176.51	55.00	30.23	27.34	43.47
283	Cys		173.42	52.71	36.29		
284	Pro			50.12			
285	Pro		177.62	65.10	32.11	27.63	50.12
286	Glu		175.21	58.17	28.95	36.54	
287	Trp	9.95	173.99	56.05	34.62		

Res	Typ	HE	C	CA	CB	CB	CG	CD
288	Thr		174.72	59.76		71.19	20.35	
289	Gly		175.77	43.86				
290	Gln	6.61 7.03	176.26	59.07		28.84	32.22	
291	Tyr	6.85	176.24	55.42		38.74		
292	Cys		174.2	54.59		34.34		
293	Thr		174.58	61.42		69.96	22.64	
294	Glu		181.49	57.60		37.00	31.73	

Chemical shifts (ppm) for hN-$1_{8\text{-}11}$ (30 mM Ca^{2+}, pH 6.1)

Res	Typ	H	N	ND	NE	HA	HB	HG	HD
294	Ala					4.34	1.36		
295	Asp	8.37	119.04			4.57	2.44 2.44		
296	Val	7.81	120.51			4.06	1.89	0.81 0.76	
297	Asp		125.18			4.77	2.90 2.35		
298	Glu	9.22	125.59			3.63	1.62 1.54	2.13 2.46	
299	Cys	7.92	112.89			4.15	3.01 3.16		
300	Gln	7.28	116.76			4.05	1.88 2.06	2.33 2.19	
301	Leu	7.4	117.1			4.1	1.56 1.53	1.6	0.78 0.78
302	Met	7.66	117.31			4.77	1.89 1.77	2.30 2.28	
303	Pro					4.31	1.80 2.26		3.62 3.62
304	Asn	8.46	115.28			4.72	2.56 2.77		
305	Ala	7.43	120.93			3.8	1.17		
306	Cys	8.21	111.81			4.71	2.45 2.25		
307	Gln	8.17	119.97			4.08	1.46 1.85		
308	Asn	8.92	113.19			3.9			
309	Gly	8.8	106.47			3.64 3.93			
310	Gly	7.28	105.66			3.37 4.32			
311	Thr	8.65	115.73			4.39	3.95	1.06	
312	Cys	8.91	128.23			5	2.78 2.99		
313	His	9.58	128.13			4.87	3.1		7.08
314	Asn	9.16	124.21			4.88	2.70 2.95		
315	Thr	8.24	116.27					1.01	
316	His					4.35	3.16		
317	Gly	8.74	112.14			3.96			
318	Gly	7.22	108.98			3.93 4.20			
319	Tyr	8.04	121.07			5.17	2.96 2.69		
320	Asn	9.39	116.48			4.88	2.69 2.68		
321	Cys	8.94	116.9			5.51	2.66 2.82		
322	Val	9.11	124.67			4.23	2.05	0.85 0.96	
323	Cys	8.98	125.84				3.38 2.66		
324	Val	8.56	121.84			4.08	2.11	1.03 1.16	
325	Asn	8.3	120.05	112.9		4.44	2.84 2.69		7.02 7.69
326	Gly	8.51	110.05			4.27 3.61			
327	Trp	8.16	121.87		127.07	5.53	3.16 3.33		
328	Thr	9.75	116.29			4.88	4.02	1.12	
329	Gly	8.03	108.21			3.93 4.81			
330	Glu	9.07	121.12			4.04	2.03 2.03	2.33 2.33	
331	Asp	8.08	113.5			5.19	2.63 2.61		
332	Cys	7.69	109.72			4.35	3.35		
333	Ser	7.7	108.38			4.51	3.96.3.96		

Res	Typ	H	N	ND	NE	HA	HB	HG	HD
335	Asn	8.96	126.73			4.04	2.43.2.84		
336	Ile	7.52	128.69			3.51	1.66	1.29 0.64	0.71
337	Asp	9.07	128.06			4.39	2.84 2.58		
338	Asp	9.3	129.09						
339	Cys	8.39	115			4.24	3.25 2.92		
340	Ala	7.04	123.68			3.99	1.33		
341	Ser	7.41	108.7			4.43	3.80 3.74		
342	Ala	6.96	123.95			3.69	0.59		
343	Ala	7.65	121.2			4.25	1.17		
344	Cys	7.46	114.39			4.44	2.64		
345	Phe	8.71	123.58			4.28	2.62 3.12		
346	His	8	120.46			4.13	2.93		
347	Gly	7.84	110.26			4.02 3.56			
348	Ala	7.94	121.56			4.7	1.36		
349	Thr	8.2	115.78			4.78	3.94	0.86	
350	Cys	9.03	127.5				2.84 2.95		
351	His	9.59	128.69			5.52			
352	Asp	9.19	124.56						
353	Arg	8.26	124.4			4.89	1.63	1.53	
354	Val	8.78	125.7						
355	Ala	9.5	132.9			4.41	1.24		
356	Ser	7.24	105.87			4.36	3.78		
357	Phe	7.64	112.2			5.21			
358	Tyr	9.04	118.94			4.91	2.74 3.05		
359	Cys	9.04	124.23			4.91	2.61 3.07		
360	Glu	9.02	125.27			4.37	1.89 1.89	2.38 2.38	
361	Cys	8.9	125.32			4.97	2.99 2.67		
362	Pro					4.68	1.98 2.34		
363	His	8.69	118.54			4.28	3.07 3.23		
364	Gly	9	113.15			3.30 4.30			
365	Arg	8.05	120.24			5.26		1.7	
366	Thr	8.8	114.41					0.86	
367	Gly	8.03	108.17						
368	Leu	9.08	122.66			4.14			0.92
369	Leu					3.8	1.51		
370	Cys	8.19	112.94			4.14	3.19		
371	His	8.95	115.51						
372	Leu	7.88	121.22			4.4			0.58 1.10
373	Asn	8.47	119.06			4.7	2.54		
374	Asp	7.82	120.79			4.74	2.52		
375	Ala	8.49	130.82			4.15	1.39		
376	Cys	8.4	113.78			4.32	3.08 3.00		
377	Ile	6.94	120			3.89	1.82	1.17 0.84	0.71
378	Ser	7.68	113.73			4.44	3.75 3.82		
379	Asn	7.83	117.09	108.86		4.55	2.87 2.95		6.75 7.44
380	Pro					4.52	1.94		
381	Cys	8.12	119.38			4.72	3.14 2.71		
382	Asn	8.77	119.37			4.48	2.72		
383	Glu	8.62	122.22			3.99	1.97 1.90	2.22 2.22	
384	Gly	8.7	112.8			3.54 4.17			
385	Ser	7.7	113.95			4.71	3.73 3.71		

Res	Typ	H	N	ND	NE	HA	HB	HG		HD	
386	Asn	8.98	120.69			4.83	2.70 2.72				
387	Cys	8.61	121.87			5.41	2.98 2.83				
388	Asp	8.99	129.28			4.93	2.49 2.49				
389	Thr	8.69	116.96			4.31	3.75		0.95		
390	Asn	9.28	127.99							6.91	7.52
391	Pro					4.3	2.06 2.30				
392	Val	8.13	116.66			3.92	1.98	0.97 0.90			
393	Asn	6.92	111.96	109		4.83	3.15 2.64			6.76	7.44
394	Gly	8.48	108.44			3.59 4.03					
395	Lys	7.99	119.09			4.28	1.77 1.77	1.27 1.26		1.60	1.61
396	Ala	8.42	121.68			4.49	1.21				
397	Ile	9.21	123.01			4.25	1.66	1.48 0.85			0.78
398	Cys	8.73	126.3			5.68	2.74 3.04				
399	Thr	8.89	116.65			4.68	4.15		1.22		
400	Cys	8.88	120.06			5.33	2.65 2.98				
401	Pro					4.46	2.40 1.92				3.91
402	Ser	8.35	114.48			4.15	4.16 3.92				
403	Gly	8.67	111.09			3.34 4.31					
404	Tyr	7.92	116.99			5.34	2.56 2.88				6.7
405	Thr	9.29	113.83			4.8	3.85		0.92		
406	Gly	7.99	107.62			4.75					
407	Pro					4.31	2.31 1.95	1.97 2.06		3.64	3.64
408	Ala	8.68	121.72			4.5	1.33				
409	Cys	8.28	112.04			4.02	2.992.91				
410	Ser	7.7	108.49			4.52	4.04 3.63				
411	Gln	8.57	122.13			4.61	1.92 1.91	2.32 2.32			
412	Asp	8.79	123.61			4.35	2.60 2.32				
413	Val	7.73	126.86			3.61	1.53	0.63 0.61			
414	Asp	8.69	126.05			4.91	2.39 2.82				
415	Glu	9.58	125.33			3.48	1.09 1.42	2.53 1.97			
416	Cys	8	112.48			4.36	3.27 3.13				
417	Ser	7.33	114.37			4.64	3.82 3.66				
418	Leu	7.35	121.32			4.23	1.55 1.50		1.59		0.7
419	Gly	8.14	107.83			3.93 3.61					
420	Ala	7.94	125.38			4.23	1.19				
421	Asn	8.51	122.79			4.75	2.25 2.83				
422	Pro							1.90 2.05			
422	Pro					4.3	1.821.69			3.62	3.61
423	Cys	7.99	118.66			4.4	2.30 2.79				
424	Glu	8.46	121.06			3.97	1.44 1.64	2.22 2.09			
425	His	8.78	113.01			3.88	2.36 2.27				
426	Ala	8.77	116.61			3.76	1.41				
427	Gly	7.15	102.45			3.31 4.09	0				
428	Lys	8.31	120.97			4.39	1.83 1.84	1.43 1.28		1.67	1.67
429	Cys	8.76	127.99			5.02	3.07 2.90				
430	Ile	9.69	131.98			4.25	01/01/88	1.41 0.79	1.19		0.76
431	Asn	9.1	126.88			4.87	2.58 2.95				
432	Thr	7.87	116.98			4.63	4.22		1		
433	Leu	8.6	123.32			4.02	1.57 1.72		1.63	0.98	0.92
434	Gly	9.14	118.95			2.92 4.17					
435	Ser	6.6	111.98			4.31	3.373.82				

Res	Typ	H	N	ND	NE	HA	HB	HG	HD
436	Phe	8.94	116.69			5.1	3.03 3.62		
437	Glu	9.81	119.56			4.61	1.99 1.84	2.07.2.11	
438	Cys	8.73	116.34			5.61	2.61 2.83		
439	Gln	9.53	124.5			4.67	2.02 2.11	2.34.2.27	
440	Cys	8.93	124.81			4.75	2.71 3.26		
441	Leu	8.75	122.41			4.46	1.76 1.80	1.75	1.02 0.97
442	Gln	8.31	117.04		111.86	3.94	2.07 2.00	2.36.2.36	
443	Gly	8.49	110.69			3.94 3.36			
444	Tyr	7.65	119.01			5.03	2.73 3.05		6.31
445	Thr	9.37	113.89			4.73	4.19	1.09	
446	Gly	7.98	106.25			4.57 3.96			
447	Pro					4.36	1.92 2.33	2.10 2.05	3.63 4.18
448	Arg	9.4	117.17		119.02	5.07	1.56 2.50	1.66 1.54	3.29 3.24
449	Cys	7.88	111.99			4.14	3.08 3.53		
450	Glu	10.8	121.5			4.2	2.15 1.79		
451	Ile	8.67	128.3			4.11	1.82	0.84 1.43 1.08	0.82

Res	Typ	HE	C	Ca	Cb	CG	CD	CE
294	Ala		177.23	52.67	19.35			
295	Asp		175.62	54.61	42.59			
296	Val		174.11	61.36	34.09	20.45 21.11		
297	Asp		176.39	51.97	39.58			
298	Glu		177.42	61.12	29.28	37.81		
299	Cys		177.08	55.23	36.34			
300	Gln		177.14	57.12	29.26	33.99		
301	Leu		177.11	56.20	43.31	26.84	23.08 25.17	
302	Met	1.91	173.18	52.75	33.18	31.58		16.96
303	Pro		177.37	64.18	31.89		50.57	
304	Asn		174.41	52.35	38.23			
305	Ala		178.18	55.23	19.02			
306	Cys		173.68	51.94	38.98			
307	Gln		176.10	55.62	30.09			
308	Asn		176.70	54.88	35.92			
309	Gly		175.16	45.63				
310	Gly		171.70	45.78				
311	Thr		172.61	61.92	70.96	21.74		
312	Cys		173.88	57.79	38.39			
313	His		173.45	54.81	31.22			
314	Asn		175.33	54.72	39.69			
315	Thr		174.59	59.88		19.23		
316	His			58.42	28.78			
317	Gly		174.33	45.39				
318	Gly		171.97	46.38				
319	Tyr		172.78	57.36	39.47			
320	Asn		173.61	52.23	42.46			
321	Cys		174.86	53.82	43.17			
322	Val		176.54	62.32	31.93			
323	Cys		176.09	53.22	37.22			
324	Val		175.04	62.55	32.91	21.41 22.14		
325	Asn		174.83	55.41	38.61			
326	Gly		173.98	44.01				

Res	Typ	HE	C	Ca	Cb	CG		CD	CE
327	Trp	9.96	175.30	55.51	32.04				
328	Thr		173.59	59.83	70.53		19.35		
329	Gly		176.77	43.69					
330	Glu		176.56	59.74	29.42	36.45			
331	Asp		162.03	52.22	41.21				
332	Cys		174.22	55.15	36.47				
333	Ser		175.09	59.29	64.35				
334	Glu		174.25	55.72	30.51				
335	Asn		175.86	51.96	37.78				
336	Ile		175.69	63.00	38.61	29.48	16.70	14.06	
337	Asp		176.28	54.37	39.59				
338	Asp		177.11	56.97	42.38				
339	Cys		176.29	57.06	40.98				
340	Ala		178.37	55.03	18.19				
341	Ser		173.72	57.12	63.53				
342	Ala		176.88	52.80	19.12				
343	Ala		176.14	51.57	18.52				
344	Cys		174.97	54.33	38.76				
345	Phe		175.3	58.65	40.50				
346	His		174.99	58.17	28.88				
347	Gly		174.13	44.86					
348	Ala		176.50	51.29	21.60				
349	Thr		172.76	62.05	70.48		20.92		
350	Cys		173.37	57.53	41.12				
351	His		173.73	54.53					
352	Asp		175.56	56.89	43.19				
353	Arg		176.10	54.42	30.58				
354	Val			65.07	29.92				
355	Ala			52.52	16.89				
356	Ser		170.77	57.94	63.54				
357	Phe		173.09	55.89	42.43				
358	Tyr		173.79	56.30	41.75				
359	Cys		171.44	52.83	40.63				
360	Glu			54.86	30.75				
361	Cys			51.79					
362	Pro		175.32	62.28	32.23				
363	His		176.36	58.49	29.16				
364	Gly		174.15	45.28					
365	Arg		174.23	53.11					
366	Thr		172.85	60.30	69.68		21.49		
367	Gly		176.65	43.51					
368	Leu		177.58	58.96	41.28			24.71	
369	Leu			53.35	41.29				
370	Cys		174.50	55.56	36.06				
371	His		174.38	56.82	31.22				
372	Leu		176.79	52.85	43.59			25.88	
373	Asn		174.97	53.27	38.58				
374	Asp		176.39	53.73	40.79				
375	Ala		178.55	53.55	19.57				
376	Cys		176.51	55.84	41.14				
377	Ile		176.59	63.55	37.66	27.64	17.67	12/01/00	

Res	Typ	HE	C	Ca	Cb	CG	CD	CE
378	Ser		173.67	57.51	62.86			
379	Asn		172.72	53.34	38.10			
380	Pro		177.33	64.00	32.25			
381	Cys		174.61	52.92	38.27			
382	Asn		175.52	53.41	39.03			
383	Glu		176.87	58.52	29.31	38.88		
384	Gly		174.26	44.96				
385	Ser		173.89	58.52	65.16			
386	Asn		173.31	52.76	40.20			
387	Cys		173.40	56.51	41.99			
388	Asp		174.21	53.58	44.74			
389	Thr		174.00	63.15	69.91	21.05		
390	Asn		174.63	51.24				
391	Pro		177.19	64.06	32.26			
392	Val		176.89	64.48	32.12	20.94 21.88		
393	Asn		176.82	51.87	39.78			
394	Gly		173.88	46.35				
395	Lys	2.95 2.95	175.73	55.97	33.53	25.69	29.00	42.32
396	Ala		177.69	52.46	19.26			
397	Ile		175.38	60.54	40.26	17.64 27.03	13.37	
398	Cys		175.27	52.68	40.15			
399	Thr		174.29	60.70	71.76	22.16		
400	Cys		173.79	51.57	34.62			
401	Pro		175.30	63.15	32.09		50.81	
402	Ser		175.32	60.32	63.37			
403	Gly		173.76	44.95				
404	Tyr	6.8	174.65	57.51	42.79			
405	Thr		173.26	59.92	69.77	18.4		
406	Gly		173.75	44.56				
407	Pro		177.71	64.82	31.52	27.34	49.90	
408	Ala		176.33	51.60	18.26			
409	Cys		174.45	54.62	35.43			
410	Ser		174.77	58.12	65.24			
411	Gln		174.87	54.51	29.99	33.76		
412	Asp		176.24	55.05	43.52			
413	Val		173.86	62.45	33.53	20.73		
414	Asp		176.85	50.97	38.90			
415	Glu		177.21	61.12	28.18	38.22		
416	Cys		176.45	53.10	35.67			
417	Ser		174.20	57.82	64.34			
418	Leu		177.27	55.51	42.40	26.26	22.94	
419	Gly		173.07	45.94				
420	Ala		176.99	51.51	19.41			
421	Asn		174.79	51.10				
422	Pro		176.33	63.43	31.90	27.22		
422	Pro						50.51	
423	Cys		173.61	53.39	39.24			
424	Glu		176.19	55.99	30.23	36.06		
425	His		175.33	57.48	25.82			
426	Ala		176.45	53.02	16.87			
427	Gly		171.39	45.94				

Res	Typ	HE	C	Ca	Cb	CG	CD	CE
428	Lys	2.94 2.95	174.40	55.51	33.92	24.90	29.31	42.28
429	Cys		174.21	58.10	35.55			
430	Ile		174.60	60.05	39.67	17.06 26.83	12.63	
431	Asn		175.28	54.79	39.35			
432	Thr		173.88	59.59	71.23	21.74		
433	Leu		176.58	56.31	40.24	26.83	24.22 24.62	
434	Gly		174.11	46.28				
435	Ser		173.29	55.96	63.01			
436	Phe		172.49	57.36	39.77			
437	Glu		174.48	54.31	33.82	36.04		
438	Cys		174.48	53.58	44.70			
439	Gln		176.03	54.66	28.89	33.74		
440	Cys		175.50	54.18	38.33			
441	Leu		176.48	54.86	43.01	27.91	25.82 23.03	
442	Gln	6.87 7.56	176.19	57.87	28.39	33.84		
443	Gly		173.58	44.58				
444	Tyr	6.87	175.16	57.98	40.21			
445	Thr		173.58	60.42	71.72	20.89		
446	Gly		174.42	43.85				
447	Pro		178.18	65.00	31.92	27.58	49.92	
448	Arg	7.3	175.69	54.27	29.58	28.39	43.17	
449	Cys		175.02	54.76	36.24			
450	Glu		175.90	58.18	30.00			
451	Ile		180.74	63.76	38.02	17.71 27.06	12.63	