SCUOLA NORMALE SUPERIORE

PISA

and

THE INTERNATIONAL CENTRE for GENETIC ENGINEERING and BIOTECHNOLOGY (ICGEB)

TRIESTE

Intrinsic Global Disorder and Inducible Local Order in the Cytoplasmic Tail of the Notch Ligand Delta-like 4

Thesis submitted for the Degree of Doctor of Philosophy

Perfezionamento in Genetica Molecolare e Biotecnologie

Candidate**: Alfredo De Biasio**

Supervisor**: Sándor Pongor, Ph.D., D.Sc.**

Academic Year: 2007-2008

Abstract

Signaling mediated by Notch receptors and their ligands is essential in cell differentiation and morphogenesis in metazoans. As both receptors and ligands are cell-surface expressed proteins, Notch signaling is restricted to nearby interacting cells. The five human ligands of Notch receptors are all single-pass, type I transmembrane proteins consisting of an extracellular region involved in receptor binding and of a 100-150 residue intracellular tail. One of these ligands, Delta-like 4 (DLL4) is a human homologue of *Drosophila* Delta protein, and plays an important role in the development of blood vessels. The intracellular region of DLL4 (DLL4_IC) is required for receptor/ligand endocytosis, undergoes regulated intra-membrane proteolysis and, through its C-terminal PDZ binding motif, mediates the interaction of DLL4 with Dlg-1, a protein involved in the organization of cell-cell junctions. The sequence of DLL4_IC is very well conserved through evolution but does not encode any domain of known structure. Using a recombinant purified protein expressed from a codon-optimized synthetic gene, we demonstrate through various biophysical methods such as circular dichroism, size-exclusion chromatography, and NMR that DLL4_IC is globally disordered in solution, but can form inter-convertible local secondary structures in response to specific variations in the physico-chemical milieu, as well as in the presence of its target PDZ domain. Most of these conformational changes occur in the functionally relevant C-terminal segment. A computational study on the incidence and location of protein intrinsic disorder in 369 human receptors of the same transmembrane class of DLL4 provides evidence that disorder concentrates in the cytoplasmic tail of these proteins and represents a general phenomenon. In light of these findings, we propose that global disorder in the cytoplasmic tail, in concert with local pre-organization, may play a role in the function of DLL4 as well as in that of other single-pass transmembrane proteins.

List of publications

Publications on the subject of this thesis:

1. De Biasio A, Guarnaccia C, Popovic M, Uversky VN, Pintar A, Pongor S. Prevalence of intrinsic disorder in the intracellular region of single-pass transmembrane proteins: the case of the Notch Ligand Delta-4. *J Proteome Res., in press*

2. Pintar A, De Biasio A, Popovic M, Ivanova N, Pongor S. The intracellular region of Notch ligands: does the tail make the difference? *Biol Direct.* 2007;2:19.

3. Popovic M, De Biasio A, Pintar A, Pongor S. The intracellular region of the Notch ligand Jagged-1 gains partial structure upon binding to synthetic membranes. *FEBS J.* 2007;274(20):5325-36.

Acknowledgments

This thesis was prepared in the Protein Structure and Bioinformatics Group at the International Centre for Genetic Engineering and Biotechnology (ICGEB) in Trieste, Italy, and was supported by an ICGEB predoctoral fellowship.

First, I need to express my gratitude to my supervisor, Prof. Sándor Pongor, for his advice and care throughout the time I spent in his lab.

I am deeply indebted to Dr. Alessandro Pintar who supervises the studies on Notch signaling within the lab, for his day-to-day help an advice throughout the entire project, and especially for his patience in taking me through the challenging yet fascinating field of protein NMR.

I am grateful to all the people in the Protein Structure and Bioinformatics Group, for having shared hundreds of morning coffees and discussions on international politics and ethical issues.

My gratitude goes to Jovana, for sharing the burden of being hostage of the “Trieste effect”. Her presence has been central for the results achieved herein.

And, of course, I thank my parents and my sister Anna, for their hope, trust and unlimited support. This thesis is dedicated to them.

Table of Contents

1. Introduction

Notch signalling

Mechanism of the core signaling pathway

Notch mediated signal transduction controls cell fate (specification, differentiation, proliferation and survival) and is a key process in tissue patterning and morphogenesis in developing vertebrates and invertebrates [1,2]. The main players in this signaling network are Notch receptors, four members of which have been identified in humans (NTC1, NTC2, NTC3, NTC4), and their corresponding ligands, belonging to two distinct families: homologues of *Drosophila* delta protein (Delta-like 1, Delta-like 3 and Delta-like 4 [i]) and homologues of *Drosophila* Serrate, Jagged-1 and -2 (JAG1, JAG2).

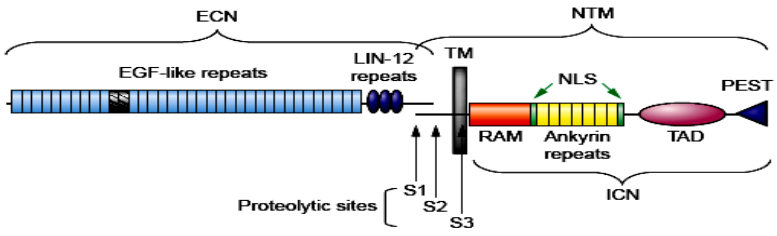

Figure 1.1 Domain organization of Notch receptors. Human Notch1 (NTC1) is shown as an example. Proteolytic cleavage by furin at site S1 produces two subunits, ECN and NTM, which remain non-covalently associated at the cell surface. EGF-like modules 11 and 12, implicated in ligand binding in *Drosophila* Notch, are shaded. S2 and S3 identify the sites of proteolytic cleavage induced upon activation by the ligand. ICN, intracellular domain of Notch; NLS, nuclear localization signal; PEST, proline, glutamate, serine, threonine rich sequence; TAD, transactivation domain; TM, transmembrane [3].

[i] Delta homologues are also abbreviated as DLL1, DLL3 and DLL4. From here on these abbreviations will be used in the text.

Notch receptors are membrane-spanning glycoproteins assembled in a non-covalent heterodimeric complex (**Figure 1.1**). The polypeptide encoded by Notch genes is proteolytically cleaved in the Golgi during the transport to the cell surface, to give an extracellular (ECN) and a transmembrane subunit (NTM). The ECN contains an array of 29-36 EGF tandem repeats, followed by three LIN-12 repeats that maintain Notch in a resting state. The intracellular region of the NTM includes a RAM domain, followed by seven ankyrin repeats, a TAD domain, and a PEST region. All Notch ligands share a similar architecture (**Figure 1.2**): a poorly characterized N-terminal region required for receptor binding, a Delta/Serrate/Lag-2 (DSL) domain, a variable number of EGF-like repeats, a trans-membrane segment, and a relatively short (~100-150 amino acids) cytoplasmic tail [4]. Jagged ligands have an additional, cysteine-rich region proximal to the trans-membrane segment.

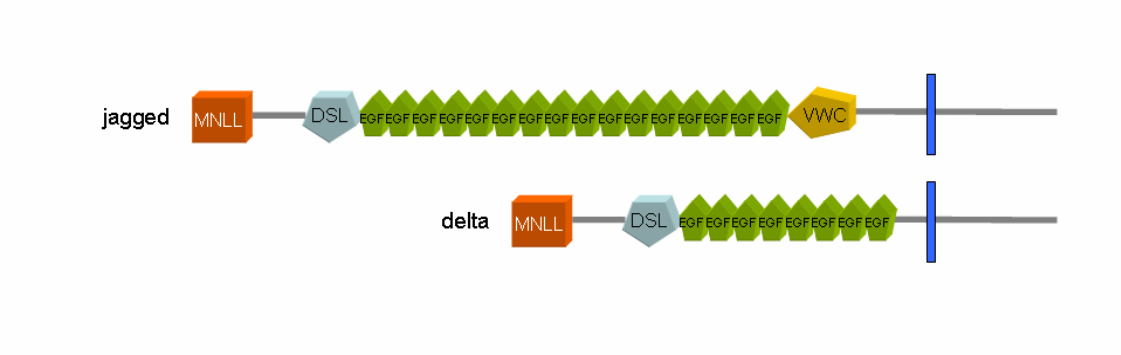

Figure 1.2. Domain architecture of Notch ligands. Typical domain organization of Notch ligands: MNLL, N-terminal domain; DSL, Delta/Serrate Ligand domain; EGF, Epidermal Growth Factor repeat; VWC, von Willebrand Factor type C domain. The transmembrane segment is shown as a blue bar. The number and type of EGF repeats can vary.

Notch signaling is initiated by receptor-ligand interactions between two distinct cells. The receptor/ligand interaction has not been characterized in detail yet. From deletion studies, it has been found that a couple of tandem EGF repeats in the receptor (EGF-11 and -12) [5] and the DSL domain in the ligand [6] are the minimal requirement for the binding to occur. In response to ligand binding, the transmembrane subunit of the receptor (NTM) is cleaved by an extracellular ADAM (A Disintegrin and Metalloproteinase) type metalloproteinase, 12 residues upstream of the membrane-spanning region. This cleavage facilitates a further cleavage of NTM, on the cytoplasmic side. This cleavage is carried

out by the presenilin/γ-secretase protease and releases the intracellular domain (ICN) from the membrane [7]. This series of controlled proteolytic events is referred to as "regulated intramembrane proteolysis" or RIP, and is a signal transduction mechanism shared with the adhesion molecules CD44 and nectin-1, the amyloid β-A4 protein, the ErbB-4 receptor tyrosine protein kinase, and others. Once translocated into the nucleus, the ICN interacts with nuclear factors that activate transcription, the main target being a transcription factor (CSL) called CBF1/RBP in mammals, Suppressor of Hairless in *Drosophila*, and LAG-1 in *C. elegans* (**Figure 1.3**).

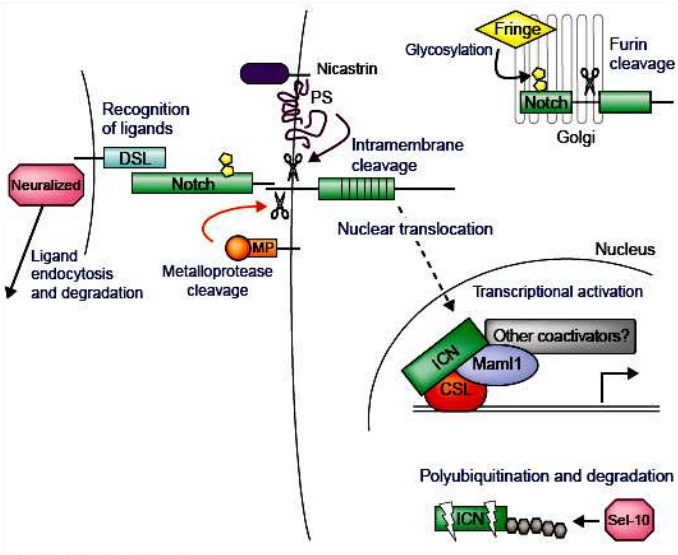

Figure 1.3. Key biochemical events in the Notch signal transduction pathway. This figure was adapted from reference 3.

Notch signaling is regulated at different levels (**Figure 1.4**): glycosylation of receptors and ligands is tuning receptor/ligand recognition [8], cytoplasmic proteins like Numb and Deltex play a role in suppressing Notch signal, E3 ubiquitin ligases regulate the level of Notch signal by targeting its components for degradation [9], and several nuclear proteins take part to the activation of transcription.

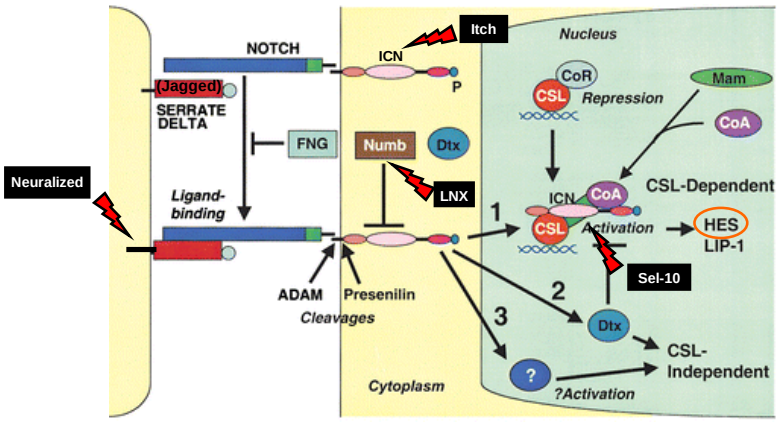

Figure 1.4. Regulation of Notch signaling. Binding of Notch receptors to ligands of the Serrate and Delta families result in successive cleavages, first in the extracellular domain by ADAM-type proteases, and then in the transmembrane domain by presenilin-dependent proteases, which release ICN and permit its translocation to the nucleus. The ability of Serrate-like ligands to activate Notch is antagonized by Fringe (FNG) glycosylases, which modify Notch extracellular domains. In the nucleus, ICN activates target gene expression by binding the transcription factor CSL, displacing corepressors (CoR), and recruiting coactivators (CoA), including mastermind (MAM) ("1"). Poorly characterized CSLindependent pathways also exist that may proceed through Dtx ("2") or unknown factors ("3"). Notch signals are negatively regulated by the cytoplasmic protein Numb, and may be positively or negatively regulated by deltex (Dtx) proteins. The figure was adapted from reference 10.

Notch signaling and cell-fate decisions

Notch signaling can have many different, if not opposite effects depending on the timing and the tissue context [11,12]. For example, while the maintenance of stem cells or progenitor cells in an undifferentiated state have been observed in the hematopoietic system and in the pancreas, terminal differentiation is induced in the skin by DLL1 or Jagged. In general, Notch signaling is acting on cell fate decisions either through lateral signaling or through inductive signaling [1]. In lateral signaling, equivalent, equipotent cells initially express both Notch receptors and their ligands, but the concentrations of these proteins start to differ between neighboring cells perhaps due to fluctuations in the steady-state expression levels. Small differences in receptor and/or ligand concentrations in cells are amplified over time, leading to cells that exclusively either express the receptors or their ligands, thus guiding the specification of the cell fate and cell differentiation. In inductive signaling, the interaction occurs between two developmentally distinct cells expressing exclusively eithet the receptor or the ligand. The fate of the bi-potential precursor cell is decided by the occurence of this interaction, while in the absence of Notch signal the precursor cell would follow another fate. The cell expressing the receptor, and therefore the recipient of the Notch signal, is induced to differentiate into a particular cell lineage.

Non redundant functions of ligands in Notch signaling

Although the molecular grounds of ligand specificity remain to be determined, *in vivo* studies suggest that each ligand exerts unique and non-redundant effects. Gene knock-out of Jagged-1 [13] or DLL1 [14], heterozygous deletion of DLL4 [15], or homozygous mutants in Jagged-2 [16] all lead to severe developmental defects and embryonic lethality in mice. Recently, impressive advances in establishing the role of DLL4 in the development of blood vessels have been achieved and several reports showed that the blockade of DLL4 signaling can inhibit tumor growth by deregulating angiogenesis [17-25] , making DLL4 a potential pharmacological target for the treatment of solid tumors [26,27]. DLL4/Notch signaling can be blocked by either a specific antibody against DLL4 that selectively

neutralizes DLL4 or a soluble DLL4 fusion protein that works by preventing Notch receptors from interacting with endogenous ligands.

Notch ligands: same dog with different tails

Notch ligands are all type I single-pass transmembrane proteins with a common architecture (**Figure 1.2**): an extracellular portion mainly made of globular domains (i.e. EGF repeats), required for receptor binding, followed by a transmembrane segment and a short intracellular region which does not encode any domain of known structure [4]. A potentially novel aspect of Notch signaling in mammals has recently emerged and underscored the central role of the intracellular region of Notch ligands in several mechanisms, such as the interaction with membrane-associated proteins, the endocytic processes that control receptor/ligand interactions, and the signaling in the ligand bearing cell, where the ligand cytoplasmic tail acts as a membrane-tethered signaling fragment. Our multiple sequence alignments showed that, within the same ligand type, the intracellular region of the five Notch ligands (DLL1, DLL3, DLL4, JAG1 and JAG2) is well conserved through evolution, while different ligand types show quite distinct cytoplasmic tails [28]. Sequence conservation within ligand types suggests that precise sequence characteristics might be required for specific patterns of post-translational modifications to take place and for specific protein-protein interactions to occur [28]. The functions mediated by the cytoplasmic tail of Notch ligands are described in detail hereafter.

The cytoplasmic tail couples Notch ligands to PDZ-containing proteins

Independent on the interaction with receptors, the cytoplasmic tail of Notch ligands couples the Notch signal transduction machinery to PDZ containing, membrane associated proteins that play a role in the organization of cell-cell junctions. Jagged-1 has been shown to interact with the unique PDZ domain of the ras-binding protein afadin (AF6) in a PDZ-dependent manner [29]. Dlg-1, the human homologue of the *Drosophila*

Discs Large protein, was identified through peptide-affinity chromatography as a binding partner for DLL1 and DLL4 [30]. It was shown that DLL1/4 can recruit Dlg-1 at cell-cell junctions, tighting cell contacts and reducing cell motility [30]. The interaction is PDZ-dependent, although it was not determined which of the three PDZ domains in Dlg-1 mediates this interaction. In similar studies, the interaction between DLL1 and members of the MAGI family (Membrane Associated Guanylate Kinases with Inverted domain arrangement) has been reported [31,32]. The interaction specifically occurs between the C-terminus of the Delta proteins and the forth PDZ domain of MAGIs. As there are over 300 human proteins containing at least one PDZ domain [4,33], it is not clear yet whether specific recognition relies on subtle differences in the PDZ domains [34], on a binding region larger then the canonical, C-terminal PDZ-binding tetrapeptide [35], or both.

Ubiquitination of the cytoplasmic tail drives endocytosis

The cytoplasmic tail of Notch ligands is also involved in ligand internalization. Although in some instances soluble forms of DSL ligands can activate Notch signals, normally an intact membrane anchored ligand is required for full activation (**Figure 1.5**). The current hypothesis is that after a receptor/ligand interaction is established, "receptor shedding" is required to expose the juxtmembrane region of the receptor to proteolytic cleavage [36-38]. Recent data suggest that receptor shedding is indeed promoted by endocytosis of the ligand/ECN complex, which physically dissociates Notch1 heterodimers before proteolysis can occur [39]. The internalization of the ligand/ECN complex is driven by mono-ubiquitination of the ligand by the E3 ubiquitin ligase. Whereas in model organisms the only apparent function of the intracellular region is to carry lysine residues that can be ubiquitinylated to trigger endocytosis [40,41], it seems that in mammals the differences in the cytoplasmic tails may underlie more specific mechanisms to control the endocytic pathways. In fact, several different E3 ubiquitin ligases that ubiquitinate Notch ligands are being identified, and different ligands are specifically recognized by different E3 ubiquitin ligases [42,43].

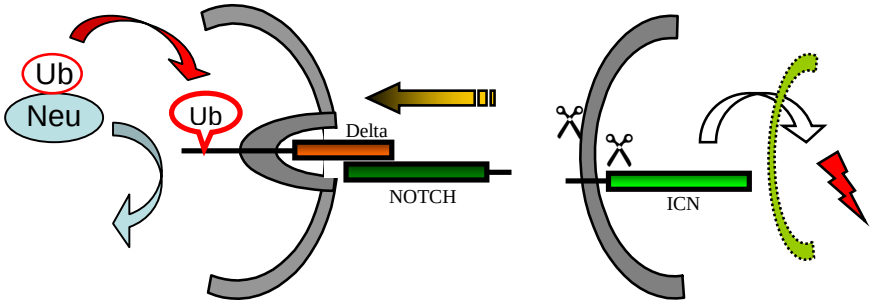

Figure 1.5. Ligand endocytosis.

The cytoplasmic tail of the ligands makes Notch signaling bi-directional

Recent reports show that Notch ligands undergo a proteolytic processing that is strikingly similar to that reported for Notch receptors (**Figure 1.6**).

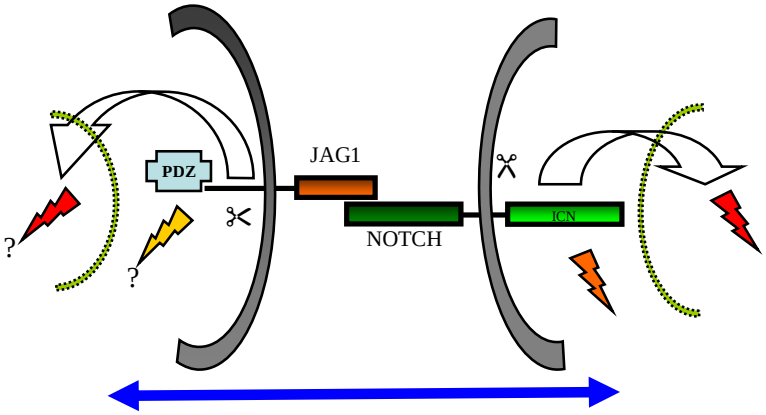

Figure 1.6. Bidirectional signaling.

Delta and Jagged undergo ADAM-mediated ectodomain processing followed by presenilin/γ-secretase-mediated intramembrane proteolysis to release signaling fragments [44-47]. The regulated intramembrane proteolysis, followed by the release from the membrane and the localization in the nucleus, suggests a possible role of the intracellular region in transcriptional regulation. In cotransfection studies, the intracellular region of

Jagged-1 was able to promote transcription of a reporter gene in COS, CHO, and HEK cells specifically through the AP1 (Activator Protein 1, p39 jun) enhancer element [46]. Activation by Jagged-1 is at odds with AP1 repression carried out by the intra-cellular domain of Notch. There is no experimental evidence, however, that the intra-cellular region of Notch ligands can bind DNA directly and, indeed, they do not contain any recognizable DNA binding motif. More probably, they function in combination with transcriptional complexes or specific transcription factors. Evidence in this direction is given by the interaction observed between the mouse DLL1 intracellular region and specific Smad transcription factors (Smad-2, -3, and -4) involved in TGF-β/activin signaling, that results in a significant enhancement in the transcription of specific genes leading to neuronal differentiation of mouse neural stem cells [48]. Moreover, co-culture experiments showed that DLL1 proteolytic processing and nuclear localization can be enhanced through interaction with Notch1 [48]. These observations strongly implies the existence of Delta signaling, which means that the Notch-Delta signaling pathway is bi-directional.

Structural biology of Notch ligands

Very little is known about the detailed molecular mechanisms involved in Notch signal transduction. Currently, structural studies on Notch signaling are mainly focused on proteins playing a role in signal transduction in the signal-receiving cell. The structure of a NL (Notch/Lin12) repeat [49], and the structure of the ligand binding region of Notch, encompassing three epidermal growth factor repeats [50], have been determined by NMR. The structure of Notch ankirin repeats have also been solved [51,52]. Of the effector proteins, the structure of CSL bound to DNA has been solved by X-ray crystallography [53]. On the other hand, little or no data has been produced concerning the structure of Notch ligands or their protein targets in the signal-sending cell. The interaction of Notch ligands with their receptors requires the DSL (Delta/Serrate Ligand) domain, but neither the structure of this domain nor the mechanism of binding has been determined. Notch signaling is sensitive to the concentration of extracellular calcium, but the effect of calcium ions on receptor and ligand structure has not been studied yet. Notch

receptor/ligand recognition is modulated by glycosylation, but the structural determinants that regulate this interaction are not known. Other post-translational modifications, like β-hydroxylation at aspartic or asparagine residues have been identified, but their role remains unclear. Also, the cytoplasmic tails of the Notch ligands, which are multi-task structural requirements needed to perform the abovementioned functions (i.e. PDZ recognition, ligand endocytosis, intra-membrane proteolysis and transcription co-activation), are still awaiting a structural characterization. Interestingly, our predictions supported by preliminary experimental results point towards a mainly disordered nature for the cytoplasmic tail of Notch ligands [28,54,55] (**Figure 1.7**), suggesting a prominent role for intrinsic disorder in the molecular mechanisms that govern the function of these proteins.

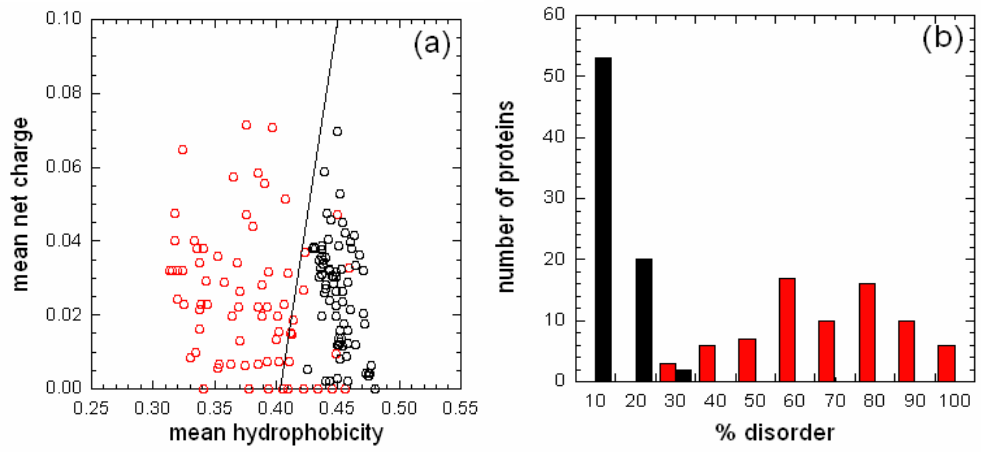

Figure 1.7. Intrinsic disorder. Disorder in the extracellular (black circles/bars) and intracellular (red circles/bars) regions of Notch ligands are shown as (a) a plot of the mean net charge v. the mean hydrophobicity and (b) as the percentage of disordered residues calculated by DisEMBL using the "hot loops" definition. In (a), the border between folded and natively unfolded proteins is drawn as a line. This figure was adapted from reference 28.

Aim of the work

The rapidly expanding experimental data underscore the importance of Notch ligands in several cellular processes. Most of the recent work has raised many issues on the role of the cytoplasmic tail of Notch ligands in bi-directional signaling, in the cross-talk with other signaling pathways, in cell-autonomous, Notch-independent signaling, and in endocytosis-mediated receptor shedding. Among the five human homologues of Notch ligands, DLL4 has proven to be a target of exceptional interest, given its central role in blood vessels development and since the blockade of the DLL4-mediated Notch signaling can inhibit tumour growth. Moreover, DLL4 couples the Notch signaling network to proteins involved in the organization of cell-cell junctions, through the interaction with the PDZ domains of Dlg-1 mediated by the ATEV motif located at C-terminus of the cytoplasmic tail. The same type of interaction observed between Notch ligand DLL1 and Dlg-1 is required to recruit Dlg-1 at the cell membrane, thereby tighting cell contacts and reducing cell motility, suggesting a similar role for DLL4 in this process. No structural characterization of the cytoplasmic tail of DLL4 has been performed so far. The cytoplasmic tail of DLL4 (DLL4_IC, 133 aminoacids) does not encode any domain of known structure and the aim of this work is to: (i) assess, through sequence analysis and biophysical studies on a recombinant protein, whether DLL4_IC encodes a new globular fold or, inversely, if it is partly or entirely disordered and (ii) identify which of the three PDZ domains of Dlg-1 interacts with DLL4_IC and study, by making use of a recombinant protein encoding the target PDZ domain, how the interaction affects the structure of DLL4_IC. Overall, this study will provide important information to describe the structural basis of DLL4 action in both the Notch-dependent and independent processes this ligand mediates in the ligand-bearing cell.

2. Results

2.1. Structural characterization of the cytoplasmic tail of DLL4 (DLL4_IC)

Sequence analysis

The amino acid sequence (**Figure 2.1**) of the cytoplasmic region of Notch ligand DLL4 (DLL4_IC) was subjected to secondary structure and disorder predictions. DLL4_IC is expected to adopt some secondary structure, as suggested by different secondary structure predictors (PSIPRED, JNet, SSpro) [56-58]. All tested methods predicted the presence of an α-helix in the N-terminal region, starting at R13, and four stretches of β conformation, two located after the α-helix and two located at the C-terminus and partially including the PDZ binding motif (**Figure 2.1**). PONDR, a disorder predictor based on neural networks trained with sequences of intrinsically disordered regions [59], predicted four disordered stretches that account for 37% of the entire sequence of DLL4_IC, whereas DisEMBL, another neural network-based predictor, predicted 55% of disordered residues, when the Hot-Loop definition of disorder was chosen (see the Materials and Methods chapter for details). IUPred [60], which is a disorder predictor that estimates the pairwise energies within an amino acid sequence, predicted 56% of disordered residues. By contrast, as expected from the globular nature of the extracellular region of DLL4 (**Figure 2.1**), the disorder prediction carried out on this region computed only 7% (PONDR) and 14% (DisEMBL) of disordered residues, while IUPred predicted no disordered regions in the extracellular domain. Overall, these predictions suggest a relevant content of protein disorder in the cytoplasmic tail of DLL4 together with the propensity of this region to adopt local secondary structures.

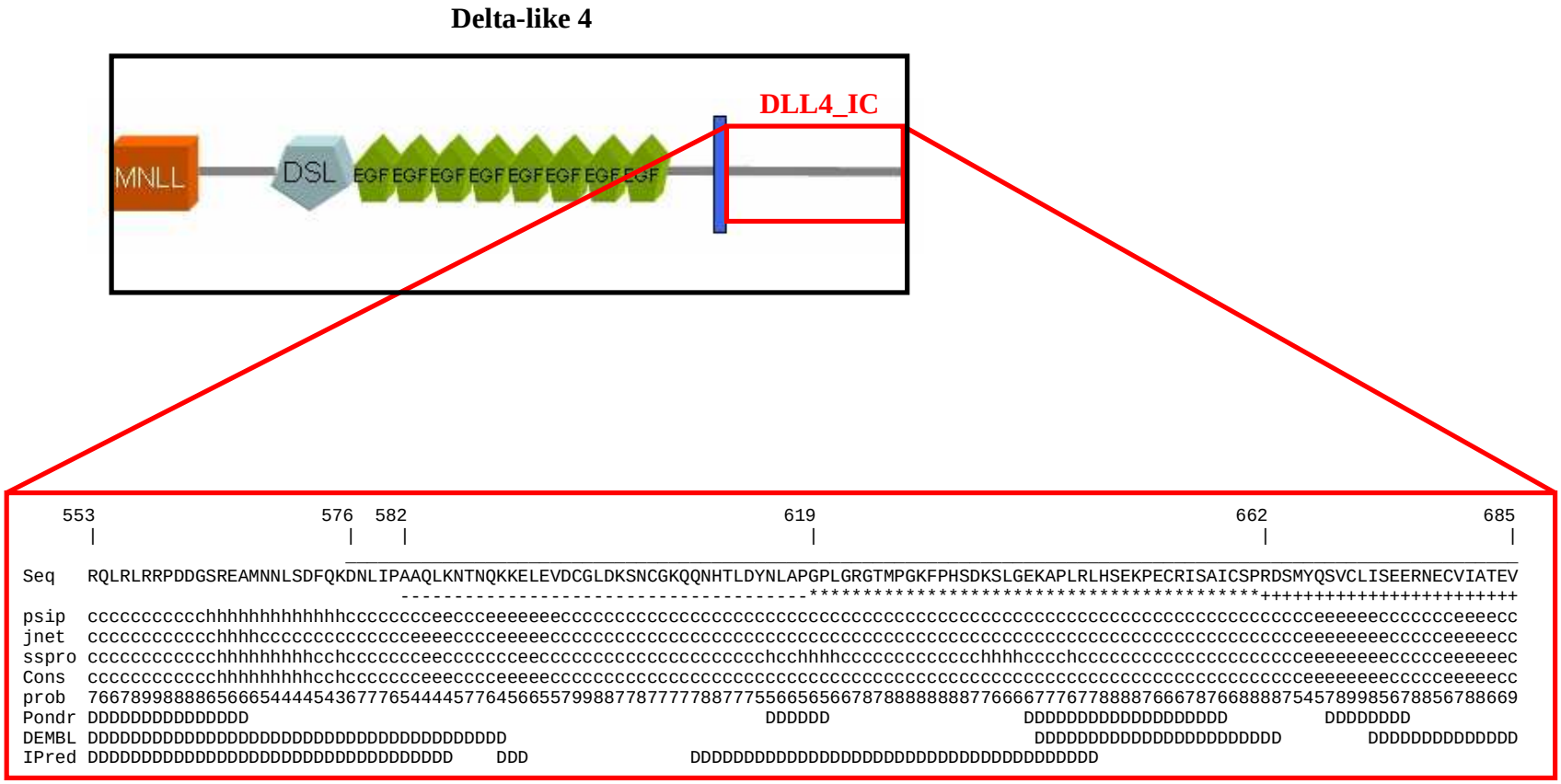

Figure 2.1. Sequence analysis and peptide design. Amino acid sequence of DLL4_IC, secondary structure and disorder predictions. Secondary structure predictions (h, helix; e, β-strand; c, coil) were obtained running PSIPRED, JNET and SSpro from the PHYRE web server (http://www.sbg.bio.ic.ac.uk/); the consensus secondary structure prediction and the score are also shown; disorder predictions (D, disordered residue) using PONDR with the VL-XT predictor, DisEMBL according to the Hot-Loop definition and IUPred with the long disorder prediction option. The solid line above the sequence (_) indicates the residues included in ΔN-DLL4_IC (res. 576-685). Segments covered by peptides are marked with (--),(P1,res.582-618),(**)(P2,res.619-661),(++)(P3,res.662-685).

Gene synthesis, protein expression and purification

The recombinant protein corresponding to the intracellular region of human DLL4 (DLL4_IC, residues 553-685 of DLL4_HUMAN, 133 amino acids) was expressed in *E. coli* from a synthetic gene designed to optimize the codon usage for heterologous expression (**Figure 2.2** and **2.3**).

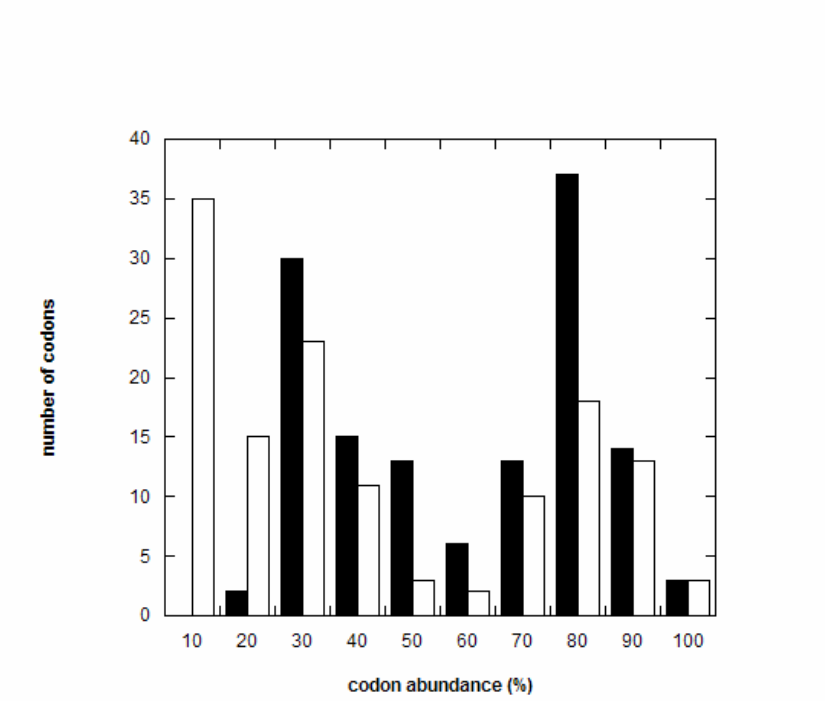

Figure 2.2. Codon usage optimization. Number of codons used in human cDNA (white) and in the synthetic gene (black) of DLL4_IC partitioned according to their relative abundance *in E. coli* Class II genes.

Despite an extensive proteolytic degradation, the final material was highly pure (>95%, as determined by RP-HPLC), and could be recovered in good yields (8 mg/L), which allowed its characterization by circular dichroism, NMR and size exclusion chromatography. The truncated protein ΔN-DLL4_IC (see below in the text) was expressed in a similar system, and recovery from inclusion bodies allowed for a single step purification by RP-HPLC.

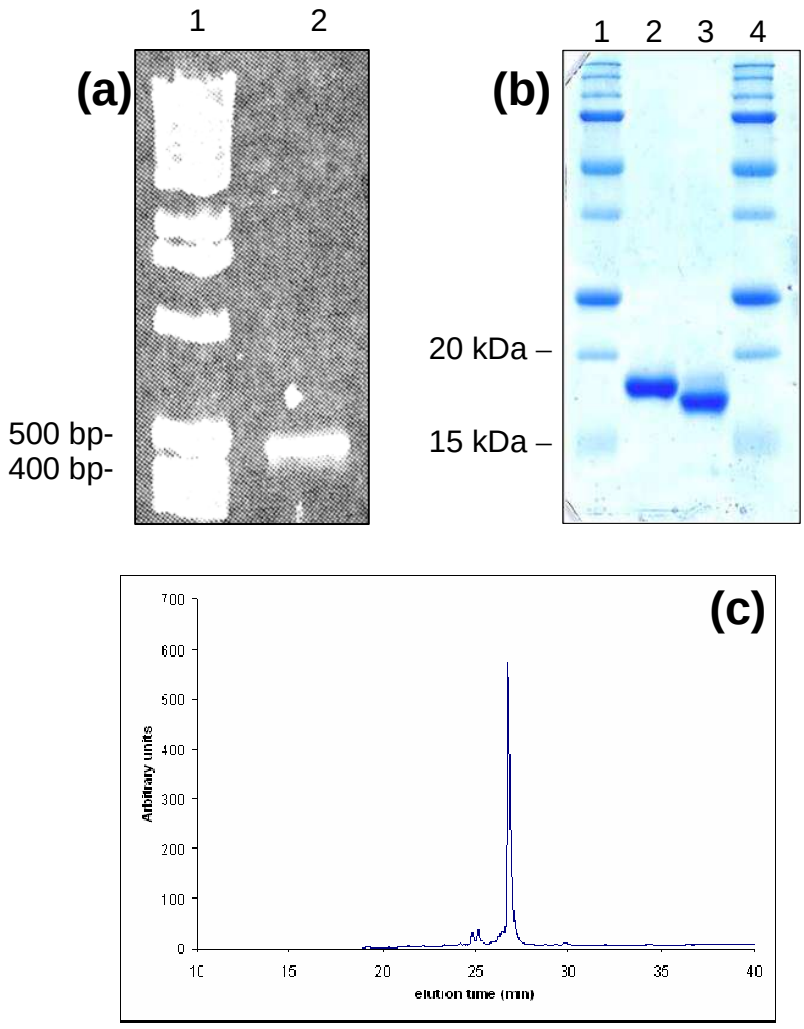

Figure 2.3. (**a**) Gene synthesis. Agarose gel (1%) of the gene assembly PCR mixture. Lane 1, 1kb DNA ladder plus; lane 2, PCR mixture. (**b**) Protein purification. Coomassie Blue stained SDS-PAGE (4-12%) of DLL4_IC before (lane 2) and after (lane 3) His_6-tag removal; lane 1 and 4, LMW markers. **(c)** Purification in native conditions. RP-HPLC analysis of His6-DLL4_IC purified by IMAC followed by ion exchange chromatography.

DLL4_IC is mainly disordered in solution

The presence of secondary structure in DLL4_IC was investigated by CD spectroscopy. The far-UV CD spectrum of DLL4_IC (**Figure 2.4**) in Tris buffer shows a strong minimum at 198 nm, which is typical of disordered proteins. The deconvolution results using CDSSTR show a high content of unordered structure (77%) and a poor residual presence of secondary structure (3% Helix, 10% Strand and 8% Turns). Very similar results were obtained from the CD spectrum of DLL4_IC purified in native conditions (**Figure 2.3c**), confirming that the purification process did not significantly affect the intrinsic conformation of DLL4_IC (**Figure 2.4**). The difference in the intensity of the negative band at 198 nm observed in the two spectra is a likely consequence of the poor accuracy in the estimation of the concentration from the measured absorbance at 280 nm. In fact, as the protein does not contain any Trp residues, this could result in more than 10% error in the computed extinction coefficient.

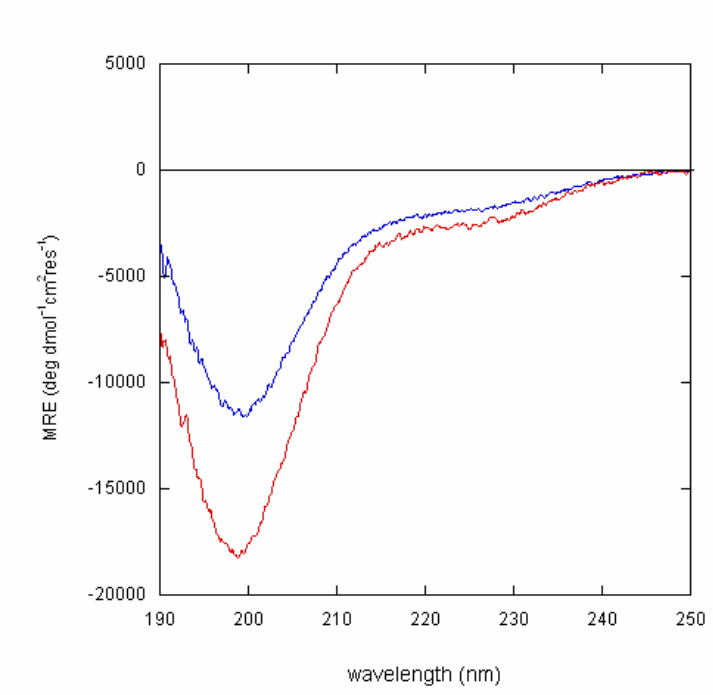

Figure 2.4. Circular dichroism. Far-UV CD spectrum of DLL4_IC (7.6 µM) in 5 mM Tris-HCl buffer, 1 mM TCEP, pH 7.5, purified in denaturing conditions (red trace) and of His6-DLL4_IC (14.1 µM) in 5 mM Phosphate buffer, 1 mM TCEP, pH 7.5, purified in native conditions (blue trace).

Also, we observed some variability in the β–sheet and turn content in different samples and over time, which might reflect the formation of protein aggregates and the appearance of a subpopulation of molecules enriched with those secondary structural elements. This could well be explained by the inherent flexibility of DLL4_IC, which is likely to make the protein structurally sensitive to slight fluctuations in the chemical environment (e.g. pH, ionic strength or temperature).

The conformation of DLL4_IC was further analyzed by NMR spectroscopy. From the ^{1}H-^{15}N HSQC spectrum of the ^{15}N-labelled protein, ~100 HN backbone resonances could be identified, which correspond to ~80% of the peaks expected, the large majority of them being clustered in a narrow region comprised between 8.0 and 8.5 ppm (**Figure 2.5a**). The average value of ^{1}HN chemical shifts (8.27 ppm) is nearly identical and the dispersion only slightly larger (σ = 0.18) compared to the values expected for a protein of the same amino acid composition and assuming random coil values [61] for all residues (8.22 ppm and σ = 0.14, respectively) (**Figure 2.5c**). The lack of chemical shift dispersion in the HN region as well as in the methyl region (data not shown) is an indicator of the lack of globular structure, and of little, if any, secondary structure. The presence of strong and sharp resonances accompanied by much weaker peaks in the ^{1}H-^{15}N HSQC spectrum, and the few peaks that could be identified in the HN-Hα region of the ^{1}H-^{15}N HSQC-TOCSY spectrum (**Figure 2.5b**) are also pointing to the presence of conformational exchange processes.

In order to better characterize the conformation of DLL4_IC in solution, we studied its hydrodynamic properties through size exclusion chromatography. DLL4_IC (15 kDa) is eluted from the size exclusion column as a peak corresponding to a 31 kDa globular protein (**Figure 2.6a**). The sharpness and symmetry of the peak indicates the presence of a single, well defined species. The elution volume of a protein from a size exclusion column correlates with its hydrodynamic properties. The hydrodynamic radius (Stokes radius, R_S) of a protein can be deduced from its apparent molecular weight (MW) as determined by size exclusion.

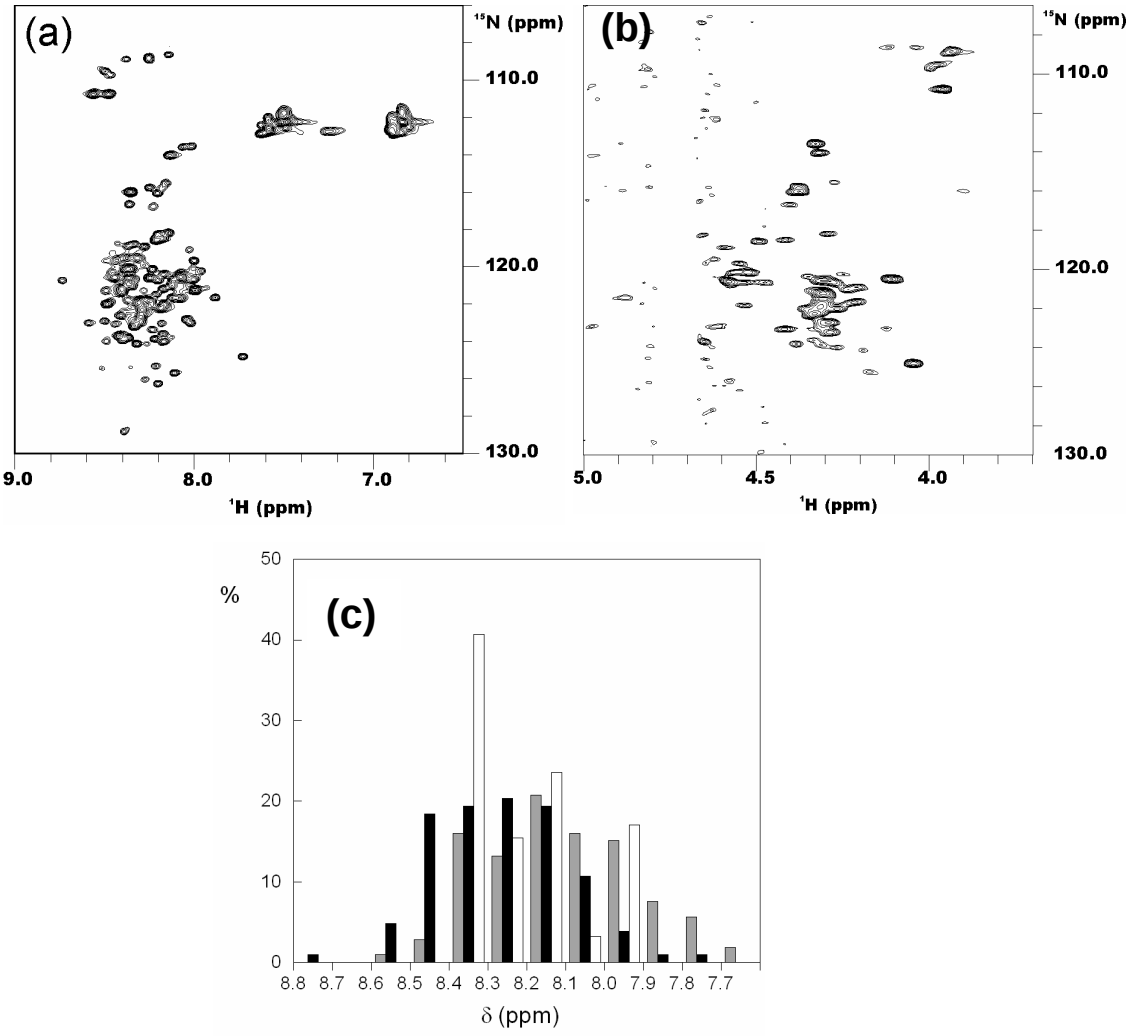

Figure 2.5. NMR. **(a)** ^{1}H-^{15}N HSQC spectra of DLL4_IC (0.5 mM) in H_2O/D_2O (90/10, v/v), 4 mM TCEP, pH 5.6, recorded at 303 K. **(b)** 1H-15N HSQC-TOCSY of DLL4_IC in H2O/D2O (90/10 v/v) containing 4 mM TCEP, 2 mM EDTA-d16, 15 mM DSS, pH 5.6, protein concentration ~0.5 mM. The spectrum was recorded at 303 K, using a 40 ms mixing time. **(c)** distribution of ^{1}HN chemical shifts of DLL4_IC in H_2O/D_2O (black bars), in the presence of SDS (grey bars) and for random coil values for a protein of the same amino acid composition (white bars).

The calculated R_S for an apparent MW of 31-kDa is 25.3 ± 0.4 Å. The theoretical radius of a monomeric protein in either a native or unfolded state (R_SN or R_SU) can be derived from its known molecular size [62]. In the case of DLL4_IC, R_SN = 19.3 ± 0.3 Å and R_SU = 35.6 ± 0.7 Å. Therefore, the large value of the Stokes radius for DLL4_IC,

experimentally determined by size exclusion (25.3 Å), intermediate between R_SN and R_SU, is consistent either with a folded, stable dimer, or with a monomeric, disordered but partially compact state. Since both CD and NMR data tend to rule out a folded globular state, we interpret size exclusion chromatography results in terms of the presence of a monomeric, disordered but partially compact state.

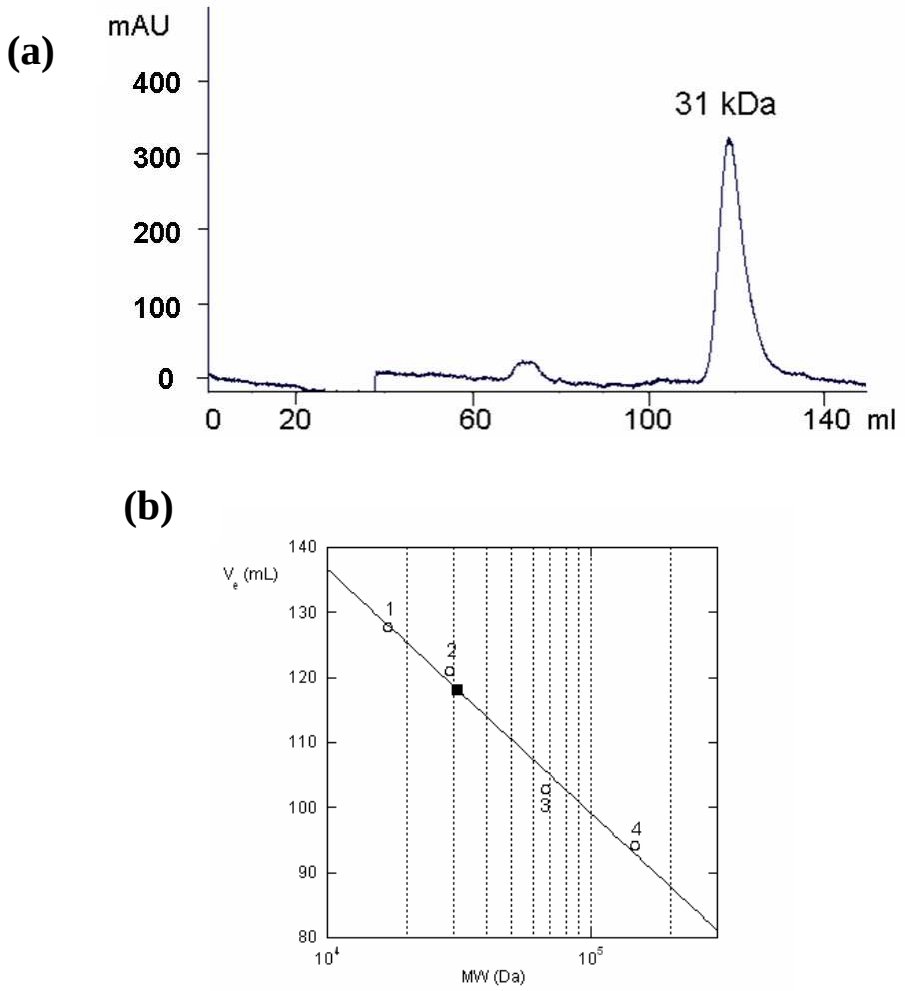

Figure 2.6.. Size exclusion chromatography. **(a)** Elution profile of DLL4_IC on a Sephacryl S-200 column (elution buffer: 50 mM Tris-HCl, 100 mM KCl, pH 7.4). The apparent molecular mass of DLL4_IC deduced from the column calibration is indicated. **(b)** Calibration standards are shown as open circles (1, lactate dehydrogenase (147 kDa); 2, bovine serum albumin (67 kDa); 3, carbonic anhydrase (29 kDa); 4, horse myoglobin (17 kDa)), DLL4_IC as a filled square (apparent MW = 31 kDa). The calibration curve (R = 0.99) is also shown. V_e = elution volume.

Taken together, CD, size exclusion chromatography and NMR data are consistent with a mainly disordered state of the protein in solution, and the presence of little or no secondary structure.

DLL4_IC displays propensity to form secondary structures

The predictions lead us to speculate that DLL4_IC secondary structure might be stabilized by agents promoting native structure or by an artificial hydrophilic/hydrophobic interface (SDS micelles). In order to test this possibility, we first analyzed the secondary structure of DLL4_IC in the presence of different concentrations of TFE, which promotes secondary structure formation by reducing the protein backbone exposure to the aqueous solvent and favoring the formation of intra-molecular hydrogen bonds [63]. Starting from a disordered conformation in aqueous solution, a significant change in the secondary structure was observed upon addition of increasing amounts of TFE. The CD spectra developed a strong ellipticity at 206 nm and a shoulder at 222 nm, characteristic of an α-helical structure, at the expense of the minimum at 198 nm (**Figure 2.7a**). The helical content increases from 3% to 17% upon TFE addition (0-20%, v/v), with a drastic change in ellipticity already between 10% and 15% TFE. Also, a significant increase in β-strand and turns structure is observed in the presence of TFE. These results confirm that DLL4_IC has the intrinsic propensity to form secondary structures, and the measured content of these is consistent with the predictions.

Similarly, at increasing concentrations of SDS, DLL4_IC undergoes a conformational change towards the α-helical structure, reaching a maximum of ~8% of α-helix at saturation (10 mM SDS, **Figure 2.7b**). Interestingly, at the same saturating SDS concentration (10 mM), the α-helical content undergoes a significant increase as the pH decreases (8% of α-helix at pH 7.5 *versus* 23% at pH 6.3, **Figure 2.7c**), while the same shift in pH only slightly increases the β-strand content of the protein alone (**Figure 2.7d**).

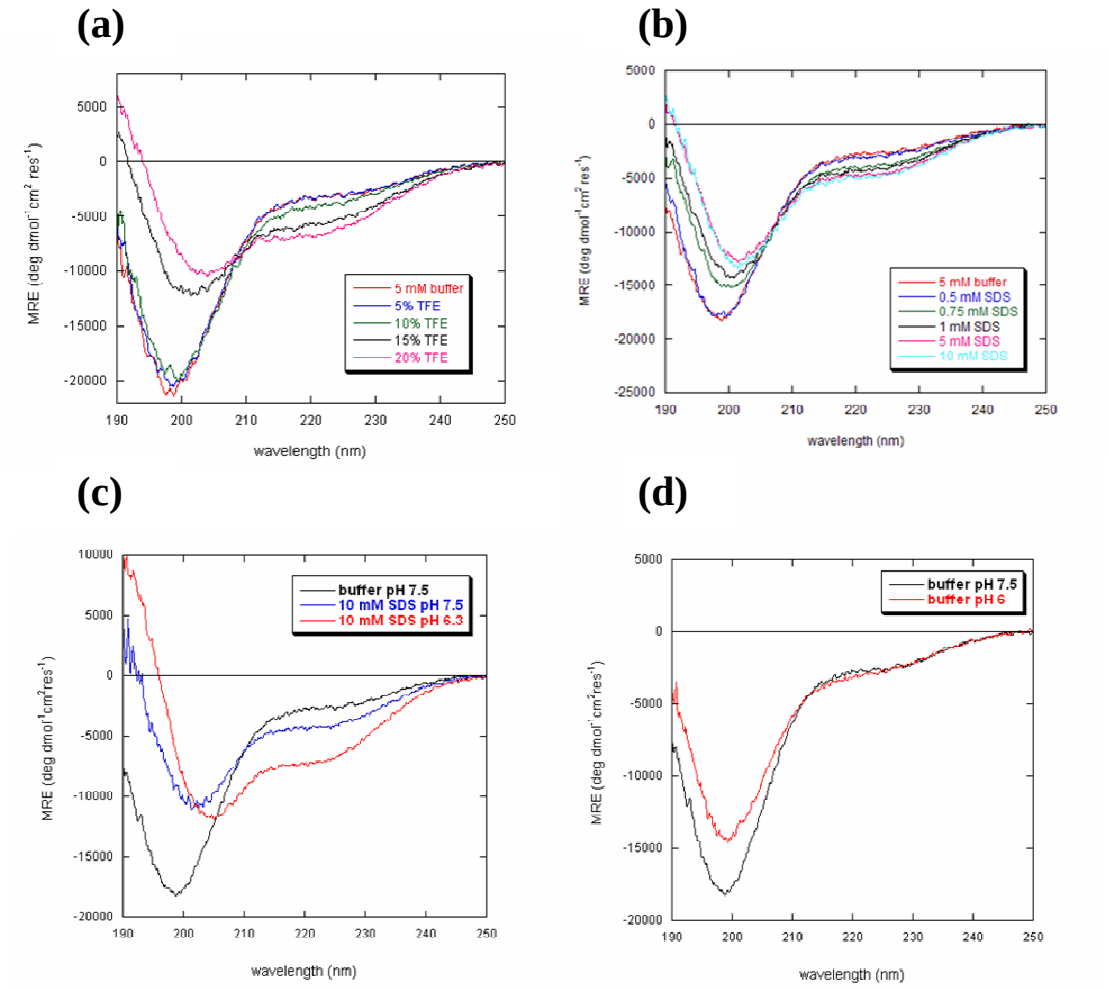

Figure 2.7. Circular dichroism. Far-UV CD spectra of DLL4_IC (7.6 µM) in 5 mM Tris-HCl buffer, 1 mM TCEP, pH 7.5, and in the presence of increasing concentrations of TFE (5, 10, 15, 20 %, v/v) (**a**) or SDS (**b**) or at different pHs in the presence (**c**) or absence (**d**) of SDS at a super-micellar concentration (10 mM)

The SDS-dependent structural change of DLL4_IC is accompanied by its association with micelles, as shown by changes in intrinsic fluorescence emission spectra and fluorescence anisotropy. An increase in tyrosine fluorescence, observed in the presence of SDS, is consistent with the interaction with the micelle surface (**Figure 2.8a**). The binding was saturable with increasing concentrations of SDS (**Figure 2.8b**). The plot of the micelle-induced fluorescence change *versus* SDS concentration was fitted with a quadratic binding equation for a two-state binding of *m* molecules of DLL4_IC to a micelle of *n* SDS molecules (see Materials and Methods). The determined value of the dissociation constant (K_d) was 2.4 µM for a complex of 4 protein molecules bound to 60

SDS molecules, a number that approximates the aggregation number for SDS (*i.e.* number of SDS molecules per micelle). More conclusive values for both the affinity and the stoichiometry of the DLL4_IC-SDS micelle complex remain to be determined. The titration with SDS reveals that SDS triggers binding below its critical micellar concentration (7-10 mM), with a saturated binding around 1 mM for 45 µM DLL4_IC, suggesting the possibility that DLL4_IC might drive the formation of SDS micelle while binding on its surface, as already seen for α-synuclein, another membrane-interacting protein [64].

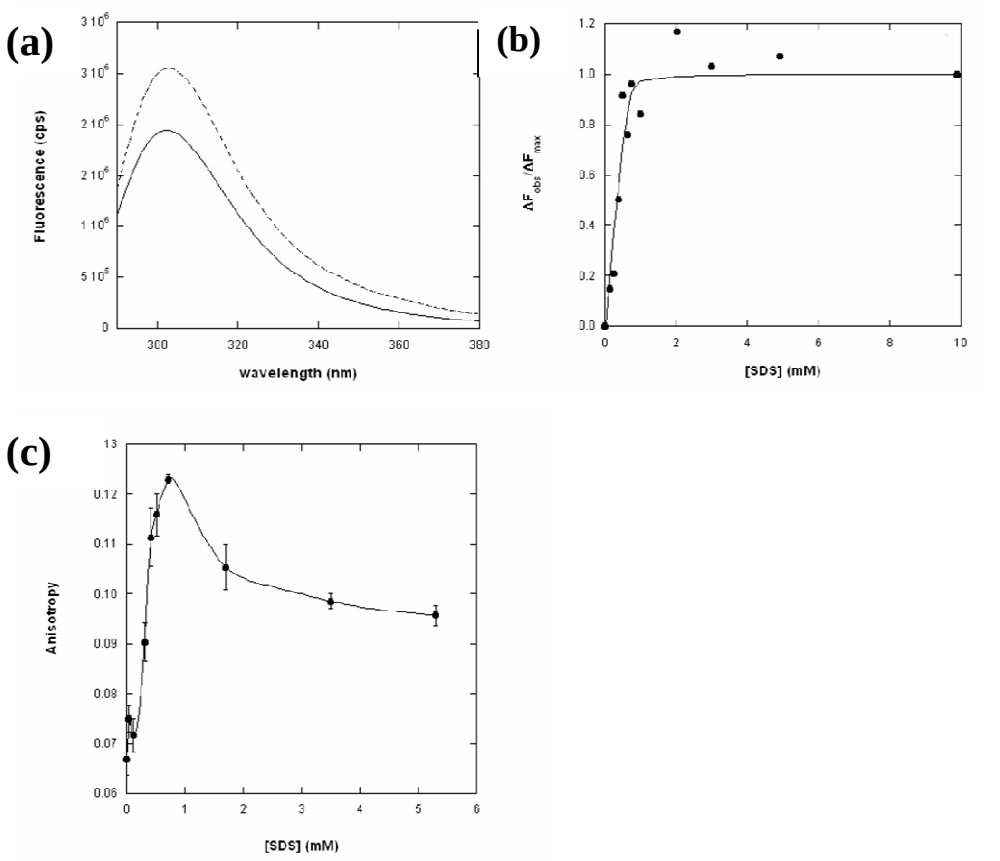

Figure 2.8.. Fluorescence spectroscopy. **(a)** Tyrosine fluorescence emission spectra of DLL4_IC (45 µM) in 5 mM Tris-HCl buffer, 1 mM TCEP, pH 7.5, in the absence (─) or presence (--) of 10 mM SDS. **(b)** Concentration dependence of DLL4_IC fluorescence increase induced by SDS. The data fitting was performed as described in Materials and Methods. **(c)** Changes in fluorescence anisotropy of DLL4_IC (45 µM) in 5 mM Tris, 100 mM NaCl, upon SDS addition (mM).

The binding of DLL4_IC to SDS micelles was further observed by changes in fluorescence anisotropy at increasing concentrations of SDS (**Figure 2.8c**). In fact, anisotropy correlates with the diffusive motions of the tyrosine fluorophore and is proportional to the size of the rotating molecule: its increase upon SDS addition is consistent with the formation of a protein-micelle complex. The binding saturates at around 1 mM SDS, confirming the results obtained from the changes in intrinsic fluorescence. The secondary structure formation in DLL4_IC observed by CD is saturated at a SDS/protein molar ratio (>100:1) significantly higher than that seen to be sufficient to saturate the protein-micelle binding as shown by intrinsic fluorescence (~20:1). This seemingly discrepant data might indicate that the protein binding to the micelle precedes the partial folding on the micelle surface, which can only be reached at a certain excess of SDS molecules.

The conformation of DLL4_IC in the presence of SDS micelles was further analyzed by NMR. The ^{1}H-^{15}N HSQC spectrum of DLL4_IC obtained at saturating concentrations of SDS is somewhat different from that of the protein alone (**Figure 2.5a** and **2.9a**). Although several resonances are still missing, probably due to overlap, HN cross-peaks appear to be of similar intensity and slightly better dispersed. Most of HN backbone resonances are still clustered in a relatively narrow region (7.7-8.4 ppm), but the average value of ^{1}HN chemical shifts (8.11 ppm) is smaller and the dispersion slightly larger (σ = 0.20) compared to the values obtained for the protein alone (**Figure 2.5c**). Moreover, ~90 cross-peaks could be counted in HN-Hα region of the ^{1}H-^{15}N HSQC-TOCSY spectrum (**Figure 2.9b**), most of them in the 4.2-4.6 ppm region. Significantly, a discrete number of non-glycine HN-Hα cross-peaks display a high field shift (< 4.2 ppm). The lack of significant chemical shift dispersion in the HN and Hα chemical shifts even in the presence of SDS micelles is an evidence of lack of tertiary structure. Also, NMR spectra suggest that DLL4_IC is conformationally restrained in the presence of SDS micelles.
Determination of secondary structure, if any, from NMR data is less straightforward. Deviations from random coil values in the chemical shifts of ^{1}Hα, $^{13}C_{\alpha}$, and ^{13}C' have been widely used to map regions with well defined secondary structure, but require residue specific sequential assignments of the backbone resonances. Recently it was

shown that also backbone ^{1}HN and ^{15}NH chemical shifts are somewhat sensitive to secondary structure [65]. The small upfield shift of these and of selected Hα resonances in the

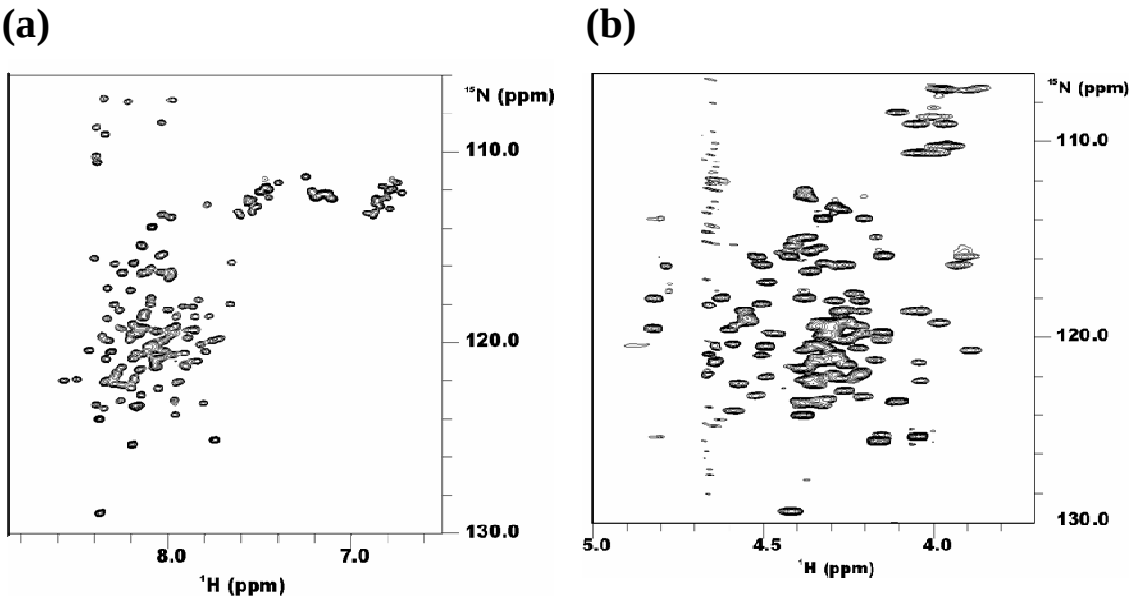

Figure 2.9. NMR. **(a)** ^{1}H-^{15}N HSQC spectra of DLL4_IC (0.5 mM) in H_2O/D_2O (90/10, v/v), 4 mM TCEP, pH 5.6, recorded at 303 K in the presence of SDS (50 mM); **(b)** 1H-15N HSQC-TOCSY of DLL4_IC in H2O/D2O (90/10 v/v), 50 mM SDS, containing 4 mM TCEP, 2 mM EDTA-d16, 15 mM DSS, pH 5.6, protein concentration ~0.5 mM. The spectrum was recorded at 303 K, using a 40 ms mixing time.

presence of SDS might then be explained in terms of partial α-helical formation, consistently with CD results. It cannot be ruled out, however, that the negatively charged head group of SDS can also contribute to the upfield shift. It has to be noticed that all NMR spectra in the presence or absence of SDS were acquired at pH 5.6, a value at which the pH-dependent secondary structures in the C-terminal region of DLL4_IC are fully formed (see below).

The N-terminal region is not required for helix formation induced by SDS

In order to test if the α–helical structure observed in the presence of TFE or SDS is located in the N-terminal region as predicted by the secondary predictions (**Figure 2.1**), we expressed and purified a truncated form of DLL4_IC (ΔN-DLL4_IC) in which the first 23 N-terminal amino acids were deleted. The CD spectrum of ΔN-DLL4_IC alone is typical of a disordered protein and is very similar to that of DLL4_IC (**Figure 2.10b**). By addition of increasing amounts of SDS, an increase in the α-helical content is observed similar to that seen with DLL4_IC titrated with SDS (data not shown). In fact, at a saturating SDS concentration (10 mM), the variation in the helical content in the two proteins is very similar (**Table 1 and Figure 2.10a and b**). Furthermore, the same pH-dependent increase in α-helix is seen in both proteins (23%). It can be concluded that the interaction of DLL4_IC with SDS micelles does not absolutely require the N-terminal stretch (res. 553-576) and that the pH-induced helical increase involves residues located elsewhere in the sequence.

On the other hand, the TFE-induced helical increase is more pronounced in the full length protein compared to ΔN-DLL4_IC (17% versus 6% in 20% TFE) (**Figure 2.10a, b** and **Table 1**). Therefore, the N-terminal region is at least partly responsible for the helical increase induced by TFE in DLL4_IC.

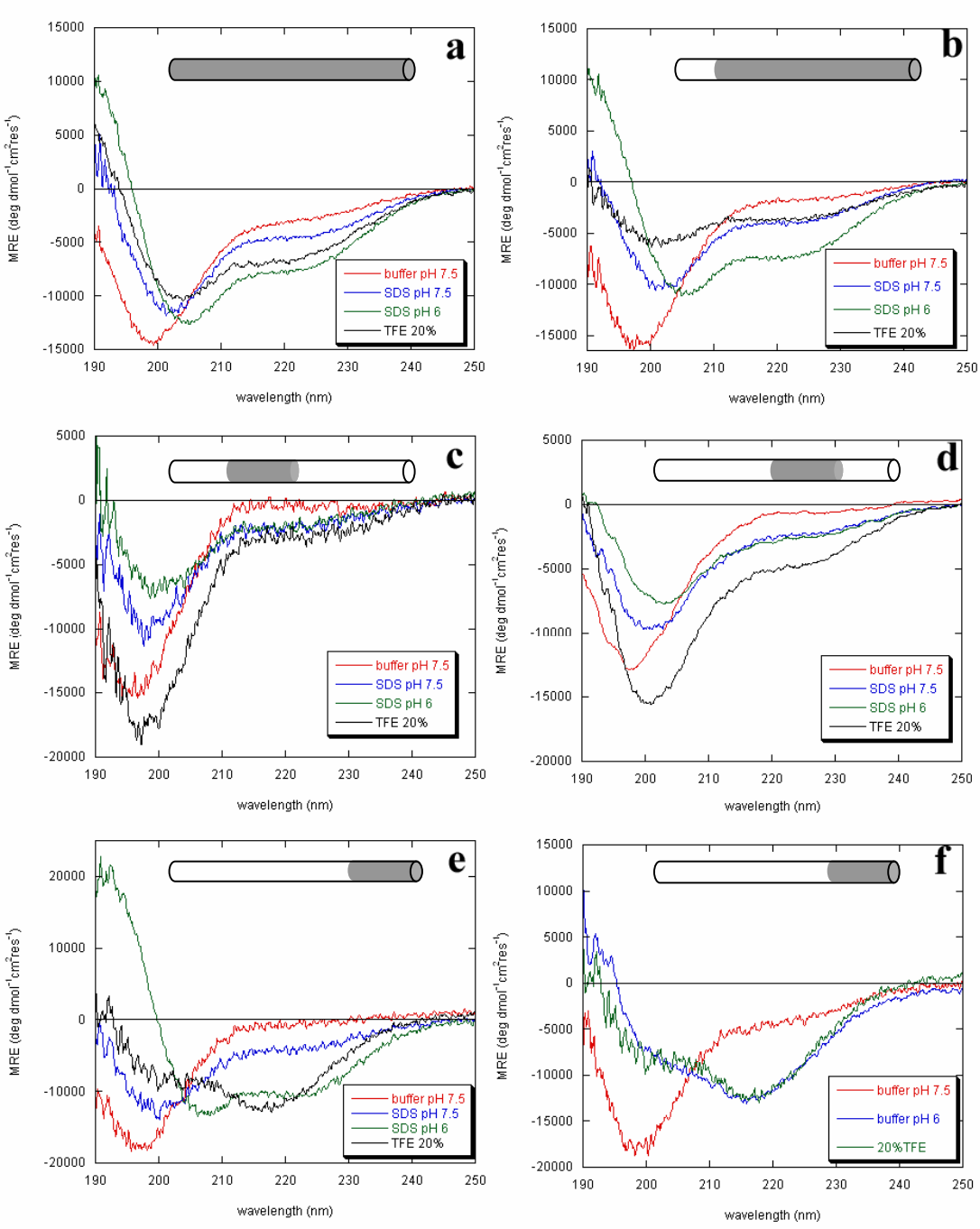

Figure 2.10.. CD of the different constructs in buffer, SDS, or TFE. Far-UV CD spectra of (**a**) DLL4_IC (7.6 µM), (**b**) ΔN-DLL4_IC (10.2 µM), (**c**) P1 (14 µM), (**d**) P2 (18.7 µM) and (**e**) P3 (17.9 µM) in 5 mM Tris-HCl buffer, 1 mM TCEP, in the presence of 10 mM SDS at pH 7.5 or 6.0, and in 20% TFE; (**f**) far-UV CD spectra of P3 in Tris buffer 5 mM (pH 7.5 or 6) or 20% TFE. The cylinder on top of each CD spectrum represents the full length DLL4_IC and the shaded area in the cylinder schematically shows the sequence covered by each construct.

		H	S	T	U	total	NRMSD
DLL4_IC	buffer pH 7.5	0.03	0.10	0.08	0.77	0.98	0.01
	buffer pH 6	0.04	0.16	0.11	0.68	0.99	0.02
	20% TFE	0.17	0.17	0.14	0.53	1.01	0.04
	10 mM SDS pH 7.5	0.08	0.16	0.13	0.62	0.99	0.02
	10 mM SDS pH 6	0.23	0.14	0.13	0.51	1.01	0.02
ΔN-DLL4_IC	buffer pH 7.5	0.03	0.19	0.12	0.66	1.00	0.03
	20% TFE	0.05	0.24	0.17	0.53	0.99	0.03
	10 mM SDS pH 7.5	0.06	0.25	0.17	0.50	0.98	0.05
	10 mM SDS pH 6	0.23	0.17	0.15	0.46	1.01	0.02
P1	buffer pH 7.5	0.01	0.14	0.09	0.74	0.98	0.02
	20% TFE	0.02	0.12	0.08	0.76	0.98	0.03
	10 mM SDS pH 7.5	0.02	0.34	0.21	0.41	0.98	0.05
	10 mM SDS pH 6	0.02	0.31	0.18	0.48	0.99	0.09
P2	buffer pH 7.5	0.03	0.14	0.10	0.73	1.00	0.02
	20% TFE	0.08	0.12	0.10	0.70	1.00	0.02
	10 mM SDS pH 7.5	0.04	0.27	0.16	0.52	0.99	0.02
	10 mM SDS pH 6	0.04	0.26	0.17	0.51	0.98	0.02
P3	buffer pH 7.5	-0.09	0.24	0.12	0.63	0.99	0.03
	buffer pH 6	0.17	0.31	0.22	0.32	1.02	0.02
	20% TFE	0.06	0.33	0.27	0.34	1.00	0.02
	10 mM SDS pH 7.5	0.05	0.16	0.11	0.66	0.98	0.05
	10 mM SDS pH 6	0.40	0.12	0.16	0.31	0.99	0.02

Table 1. Secondary structure analysis. Fraction of secondary structure (H, helix; S, strand; T, turn; U, unordered; NRMSD, normalized root mean squared deviation) calculated by CDSSTR from far-UV CD spectra.

The central region is mainly disordered

To identify the region of DLL4_IC that undergoes the pH-dependent conformational switch in the presence of SDS micelles, we synthesized and studied the conformation of three peptides (P1, P2 and P3) that cover the entire DLL4_IC sequence apart for the N-terminal region. The DLL4_IC sequence was split at proline sites, because of its helix-breaking properties, and the boundaries were selected taking into account the predictions by PONDR (**Figure 2.1** and **Figure 2.11**). According to the fraction of disordered residues in the segments relative to each peptide predicted by PONDR, the propensity to acquire structure displayed by the three peptides should have the following order:

P1>P3>P2. The conformation of the three peptides in buffer alone and in the presence of TFE and SDS was studied by far UV-CD spectroscopy and the results from the deconvolution of the spectra are summarized in **Table 1**.

P1 is disordered in buffer and its conformation is little affected by TFE (**Figure 2.10c**).

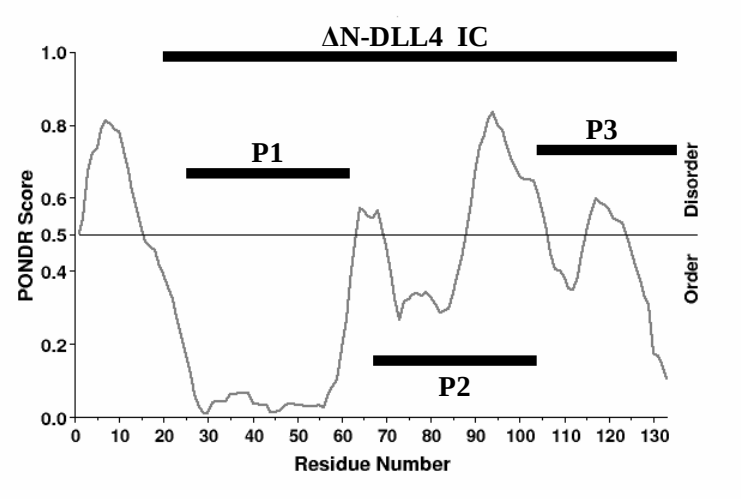

Figure 2.11. Peptide design and PONDR prediction of unstructured regions in DLL4_IC. The prediction score is plotted against the residue number. Regions with a score higher than 0.5 are considered to be disordered. The peptides are represented by the black bars above the residue number axis.

No significant structural change is observed in the presence of 10 mM SDS apart for a slight increase in strand content (15%), which is pH independent (**Figure 2.10c**).This result is in contrast with the disorder prediction, as the sequence segment that spans P1 is predicted by PONDR to be ordered.

Apart for a slightly higher helical propensity, P2 displays a conformational behavior analogous to that of P1 (**Figure 2.10d**), consistently with the high proline content of this region. These results indicate that the pH-induced helix does not form in the region covered by P1 and P2.

The C-terminal region displays structural plasticity

P3 in Tris buffer at pH 7.5 is mainly disordered (**Figure 2.10e**), with little strand content. TFE induces a relevant conformational change in the peptide, by increasing its strand and turns content up to 60%, with no increase in the helical content (**Figure 2.10e, Table 1**). This is consistent with sequence analysis results, which predict two stretches of strand at the C-terminus, partially including the PDZ binding motif (**Figure 2.1**). In the presence of 10 mM SDS at pH 7.5, a slight increase in the helical content is observed. This change is drastically enhanced when the pH is lowered to 6 (~ 40%, ~10 residues), with a concomitant decrease in β-strand content (**Figure 2.10e**). The conformational switch is fully reversed when the pH is raised back to 7.5 (data not shown). These data indicate that the pH-dependent conformational switch in P3 in the presence of SDS micelles accounts for most of the changes in the CD spectra observed for the full-length DLL4_IC, and is due to the formation of a ~10 residue α-helix in the C-terminal region (res. 662-685) of DLL4_IC. In the same conditions, a higher number of residues is calculated from CDSSTR to be helical in the full length DLL4_IC (~19 residues). However, the helical content difference estimated from the mean residue ellipticity at 222 nm is reduced, suggesting that most of the pH-induced helix forms indeed at the C-terminus. When plotting the pH dependence of the helix formation in P3 (**Figure 2.12**) in the presence of SDS using the ellipticity value at 222 nm, a steep transition at pH ~6.8 can be spotted (data not shown). Although this pH value is remainder of the pK_a of the imidazole side chain of histidines, no histidine residue is present in P3. Thus, charge neutralization of one or more of the five acidic residues (plus the carboxy-terminus) present in P3 is likely to be responsible for the pH-dependent transition. Interestingly, in 20% TFE, pH 7.5, P3 adopts the same conformation it has in Tris buffer at pH 6 (~60% strand+turns, little or no α-helix, **Figure 2.10f**).

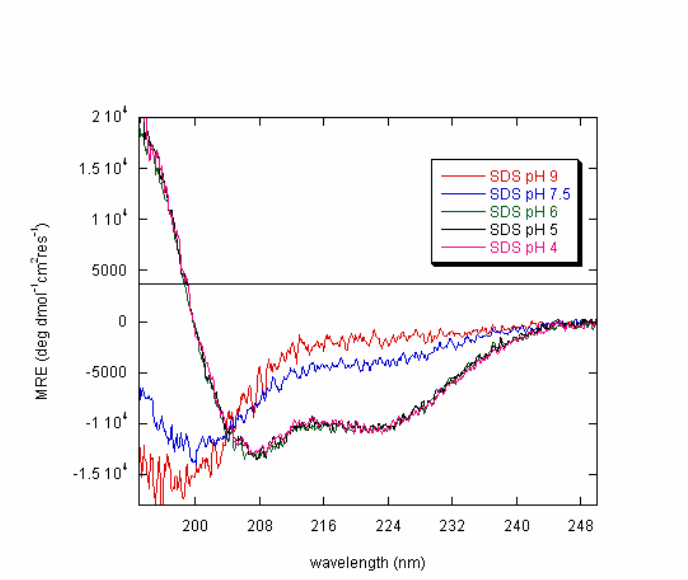

Figure 2.12.. far-UV CD spectra of P3 (17.9 µM) in 5 mM Tris-HCl buffer, 1 mM TCEP, in the presence of 10 mM SDS at the indicated pHs.

Conclusions

Based on the sequence analysis and experimental data, we can conclude that DLL4_IC is globally disordered in solution. However, it can form inter-convertible secondary structures (coil, strand and helix) in the C-terminal region that spans the last 24 aminoacids, depending on the physico-chemical environment. The plastic C-terminus is disordered in water at neutral pH but forms a strand-like structure when the pH is slightly acidic and in the presence of 20% TFE. Alternatively, at the same slightly acidic pH, it adopts a mainly helical conformation in the proximity of the hydrophilic/hydrophobic interface of SDS micelles. This C-terminal region is of functional relevance, as it includes the type I PDZ binding motif (ATEV) required for the interaction between DLL4 and its cognate protein Dlg-1 [30]. In order to better understand the structural grounds of the DLL4/Dlg-1 interaction, first we identified which of the three PDZ domains of Dlg-1 interacts with DLL4_IC *in vitro*, and then, using purified recombinant proteins and peptides, we assessed if the disordered conformation of DLL4_IC is affected either globally or locally by the interaction with its target PDZ domain.

2.2. Studying the interaction between the cytoplasmic tail of DLL4 and its target PDZ domain

PDZ domains: structure and function

PDZ domains are the most common protein interaction modules and were originally identified in the postsynaptic density protein PSD-95/SAP90 [66], *Drosophila* septate junction protein Discs-large and the epithelial tight junction protein ZO-1 [67], hence the acronym PDZ. PDZ domains are highly conserved 80-100 amino acid sequences specialized for binding the C-termini of partner proteins, generally transmembrane receptors and channel proteins, and/or other PDZ domains. Such interactions localize membrane proteins to specific subcellular domains, thereby enabling the assembly of supramolecular complexes. The role of PDZ domains in clustering and localization of proteins at the plasma membrane has relevant biological implications (*e.g.* in signaling, in ion transport, in mediating the adhesive properties of particular cells and in the formation of the specialized intercellular barriers known as tight junctions).

The structure of PDZ domains (**Figure 2.13**) includes six β-strands (βA–βF) and two α-helices (αA and αB), which fold into a six-stranded β-sandwich domain (**Figure 2.13A**). As the amino and carboxyl-termini of PDZ domains are close together, the incorporation of the domain into different multi-domain proteins is easy [68]. PDZ domains specifically recognize short carboxy terminal peptide motifs of about five residues. These sequences are often found in the cytoplasmic tails of transmembrane receptors and channels [69]. Peptide ligands bind in an extended groove between strand βB and helix αB, hence forming an additional antiparallel β-strand within the PDZ domain (**Figure 2.13**). This mechanism is referred to as β-strand addition [70]. The structure of the PDZ domain is not significantly affected by the binding to the ligand. The crystal structures of complexed and peptide-free third PDZ domain of PSD-95 are almost identical, showing RMSD between the α carbon atoms of 0.9 Å [71].

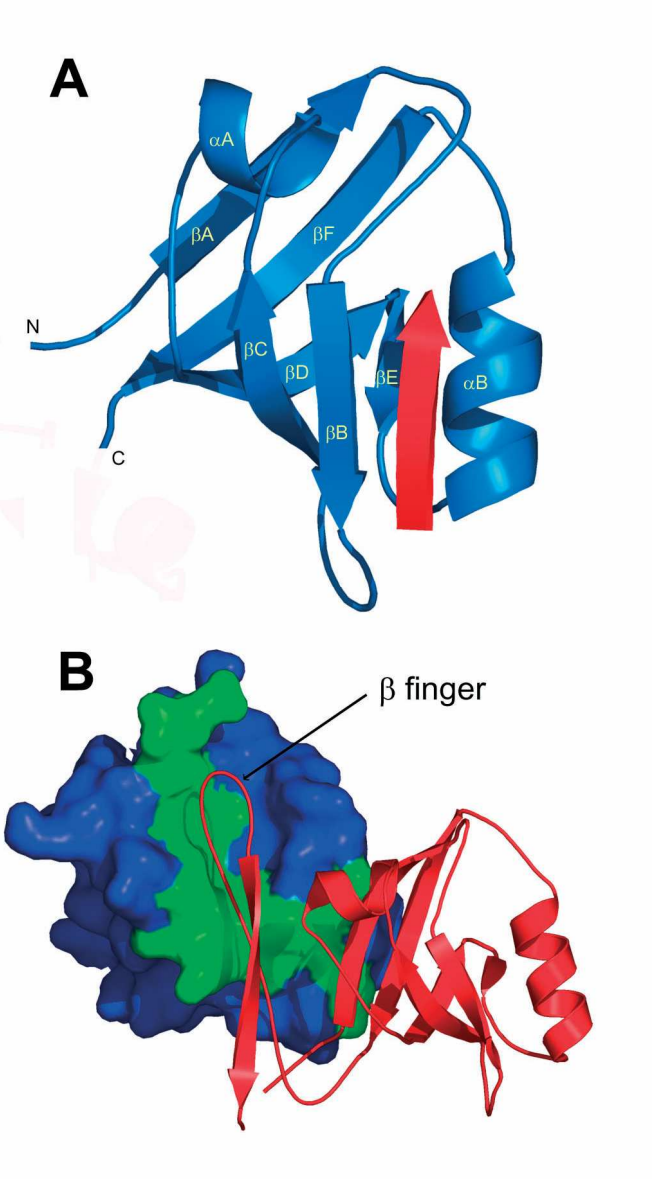

Figure 2.13. Structure of the PDZ domain bound to peptide and internal peptide motif.

A. Ribbon representation of the third PDZ domain of PSD95 (blue) with KQTSV peptide forming antiparallel β-sheet with βB strand (red arrow) (PDB code 1be9). Numering of β-strands and α-helices is shown. **B.** Complex of the syntrophin PDZ domain (shown as blue and green solvent-accessible surface representation) and nNOS PDZ domain (shown as red ribbon representation with β-finger indicated) (PDB code 1qav). This figure was adapted from reference 72.

The specificity of the interactions between PDZ domains and their ligands is imparted primarily by the sequence of the C-terminus of the ligand. Specifically, the specificity has been traditionally attributed to the last three residues of the ligand (i.e. positions P-0, P-1,

P-2, counting backwards from the terminal residue in the ligand). A classification of PDZ-binding motifs in the C-termini has been proposed, in which the consensus sequence for type I is S/T–X–Φ, and for type II is Φ–X–Φ (where Φ is any hydrophobic residue), with the corresponding PDZ-domain classified into type I or type II binding [73].

DLL4_IC interacts with the first PDZ domain of Dlg-1 *in vitro*

Dlg-1, a human homologue of the *Drosophila* Discs large tumor suppressor which is a member of the membrane-associated guanylate kinase family of molecular scaffolds, is a 97 kDa protein that bears different protein recognition domains, including a SH3 domain, a guanylate kinase homologous (GuK) region, and three PDZ domains (**Figure 2.14**).

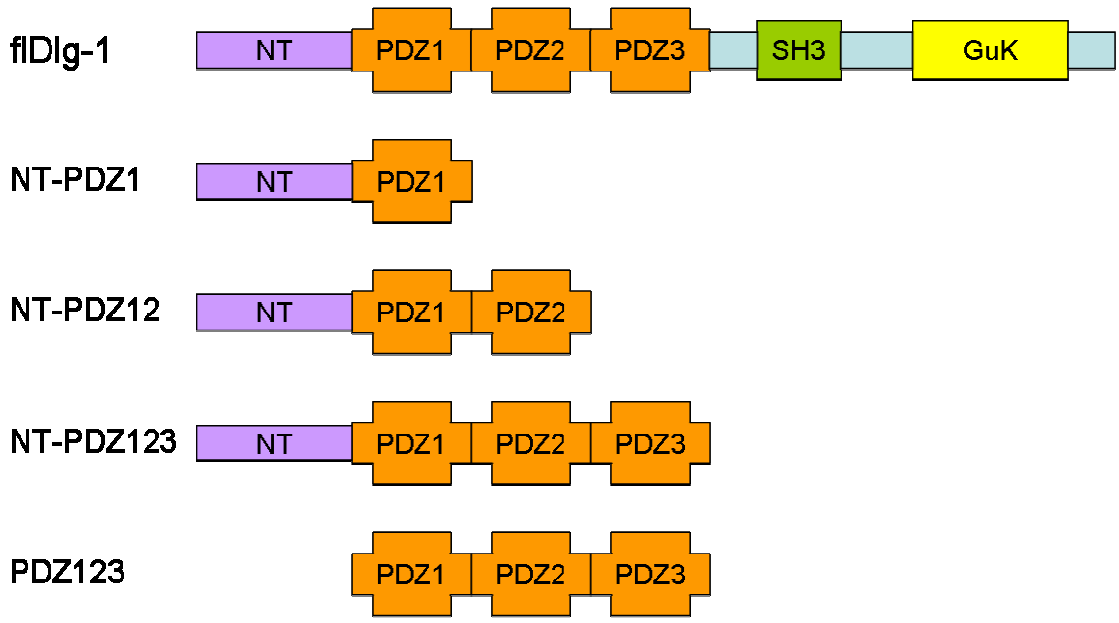

Figure 2.14. Schematic representation of full-length and deleted mutant Dlg proteins. NT refers to the amino-terminal region of the Dlg protein prior to the first PDZ domain; NT-PDZ1 mutant, aa 1–276 of the reported sequence; NT-PDZ12, aa 1–382; NT-PDZ123, aa 1–511 PDZ123, aa 186–511; SH3, Src homology 3 domain; GuK, guanylate kinase homologous region.

In order to identify which of the three PDZ domains of Dlg-1 interacts with the cytoplasmic tail of DLL4, we generated different Dlg-1 constructs, shown schematically

in **Figure 2.14**. The Dlg-1 proteins were ^{35}S radiolabeled and *in vitro*-translated with rabbit reticulocyte lysate and incubated in the presence of the purified recombinant His-tagged DLL4_IC protein previously immobilized on sepharose nickel beads. The results of this pull-down assay shown in **Figure 2.15** prove that DLL4_IC binds the full length Dlg-1 protein weakly, as a faint band could be detected when the gel was exposed to the autoradiography screen for an extended time (**Figure 2.15**, top right).

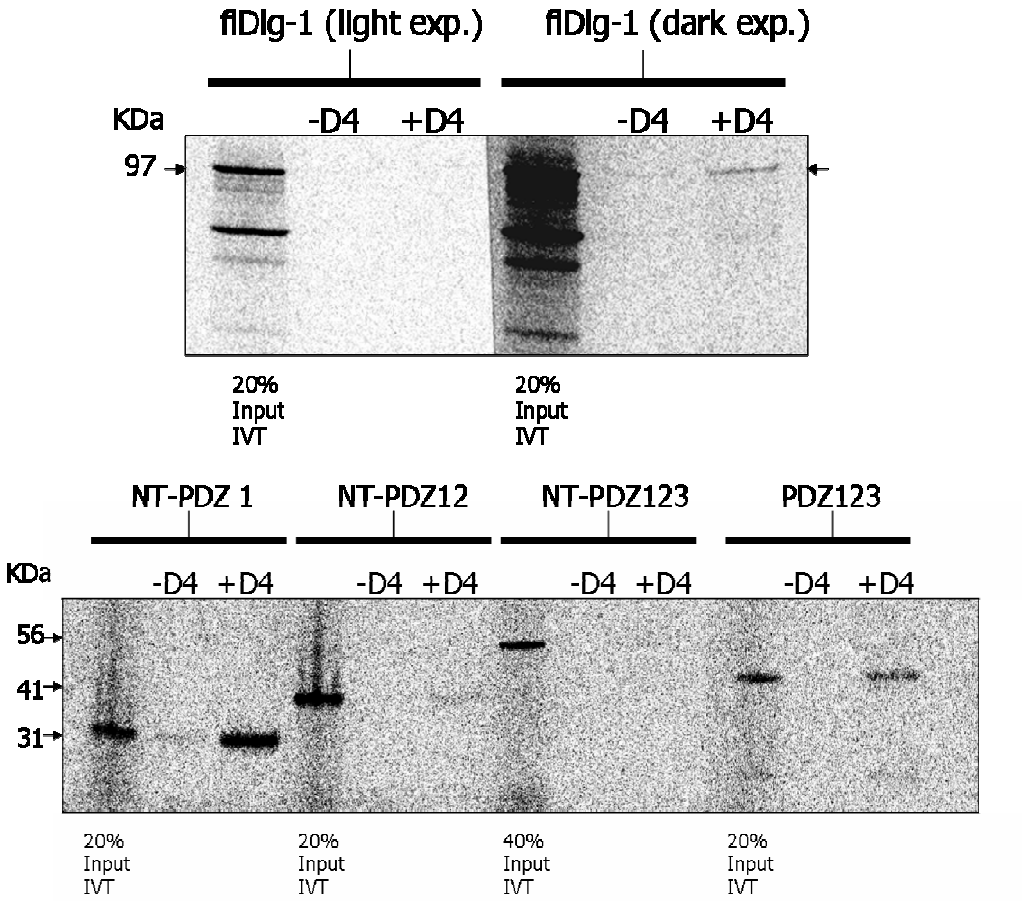

Figure 2.15. Binding of Dlg to DLL4_IC *in vitro*. ^{35}S labeled, *in vitro*-translated (IVT) Dlg-1 and the truncated mutants were incubated at 4 °C in the presence of nickel beads with (+D4) or without (-D4) His_6-DLL4_IC. After extensive washing, the Dlg-1 mutants were eluted from the beads, run on an SDS-PAGE gel (12%) and assayed by autoradiography.

However, a much stronger binding is seen between DLL4_IC and the Dlg-1 deletion mutant including the first PDZ and the N-terminal region only (NT-PDZ1, **Figure 2.15,** lower panel). The equally strong binding observed between DLL4_IC and the Dlg-1 mutant including all three PDZ domains only (PDZ123) confirms that the first PDZ is a sufficient recognition determinant in Dlg-1 that binds DLL4_IC *in vitro*. Surprisingly, the NT-PDZ12 and NT-PDZ123 mutants do not show a significant interaction with DLL4_IC. A possible explanation of this is that either the PDZ domains of those Dlg-1 mutants are not correctly folded in the *in vitro*-translation mixture, or the binding site in the first PDZ domain is somehow masked by the neighboring PDZ domains when the N-terminal region is present. The latter behavior could be related to the oligomeric state of the Dlg-1 protein, as the N-terminal region contains an L27 domain known to form a tetrameric complex in solution [74]. These factors could also be at the origin of the weak interaction observed with the full length Dlg-1 protein. Although these data do not give information on the different affinity of each of the three PDZ domains of Dlg-1 towards DLL4_IC, they show that the first PDZ domain (PDZ1) binds DLL4_IC *in vitro* (**Figure 2.16**). We therefore proceeded with the expression and purification of a recombinant protein encoding PDZ1.

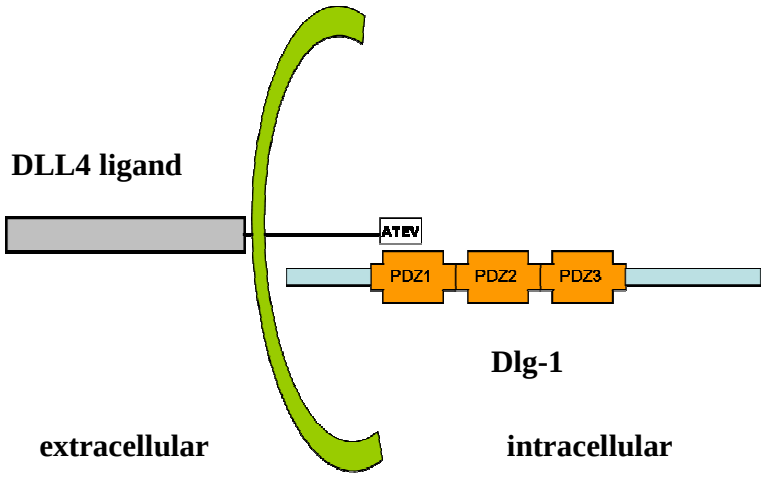

Figure 2.16. Pictorial representation of the interaction between the intracellular domain of Delta-4 and the first PDZ domain of Dlg-1.

Expression and purification of the first PDZ domain of Dlg-1 (PDZ1)

The recombinant protein corresponding to the first PDZ domain of rat Dlg-1 (PDZ1, residues 221-311 of DLG1_RAT) was expressed in *E. coli* with an hexahistidine tag at the C-terminus. The protein was purified to homogeinity (10 mg/L) with a single IMAC step followed by a desalting step (**Figure 2.17a**). The correct folding of the protein was assessed by far UV-CD spectroscopy (**Figure 2.17b**). The deconvolution of the spectrum using CDSSTR computed 19% helical and 26% beta sheet residues, while the Stride assignment from the PDB file of the NMR solved structure reports 16% and 26% of those secondary structural elements, respectively [75,76].

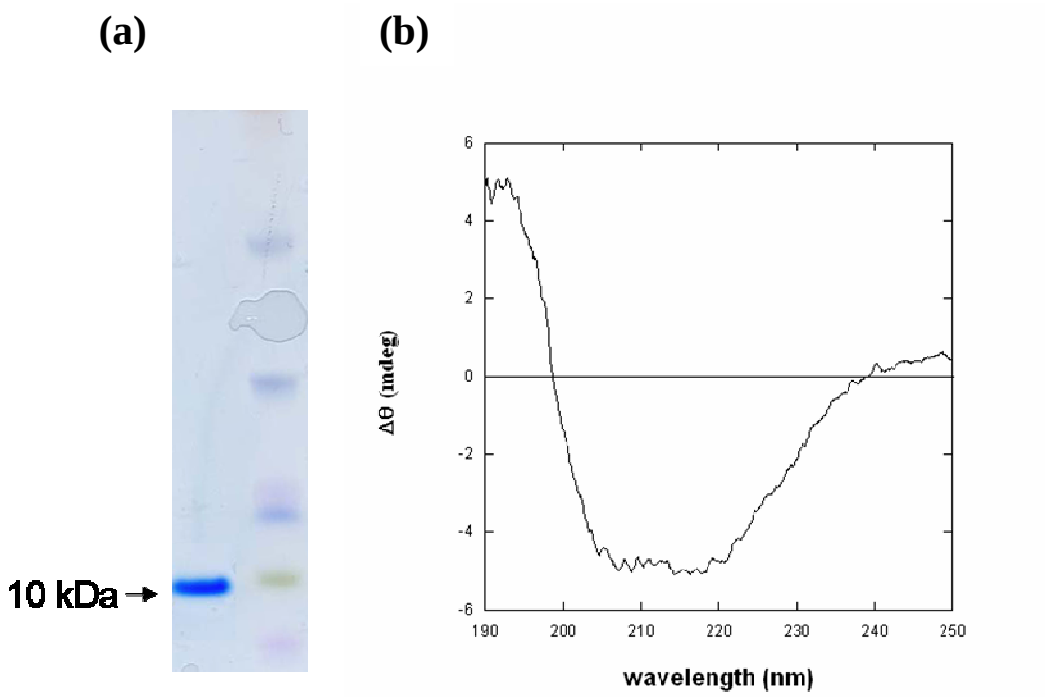

Figure 2.17 SDS-PAGE (**a**) of PDZ1 expressed in E. coli and purified, and (**b**) CD spectrum of the purified PDZ1.

The C-terminus of DLL4 interacts with PDZ1

The *in vivo* interaction between DLL4 and Dlg-1 requires the tetrapeptide (ATEV) located at the extreme C-terminus of DLL4 [30] (**Figure 2.16**). Using NMR spectroscopy, we showed that the C-terminal region spanning the last 24 amino acids of DLL4 (P3 peptide) interacts with the first PDZ domain of Dlg-1. The assay was performed by recording the ^{1}H-^{15}N heteronuclear single-quantum correlation (HSQC) spectra of the purified and uniformely ^{15}N labeled PDZ1 protein on addition of substechiometric amounts of unlabeled P3 peptide. In the course of the titration, the majority of the resonances shifted (red peaks in **Figure 2.18**) in the spectra, whereas some peaks disappeared and a small new set of peaks appeared. The resonance shift increased proportionally to the increasing amounts of peptide, suggesting a fast exchange on the NMR time scale between the free and peptide-bound form of the PDZ.

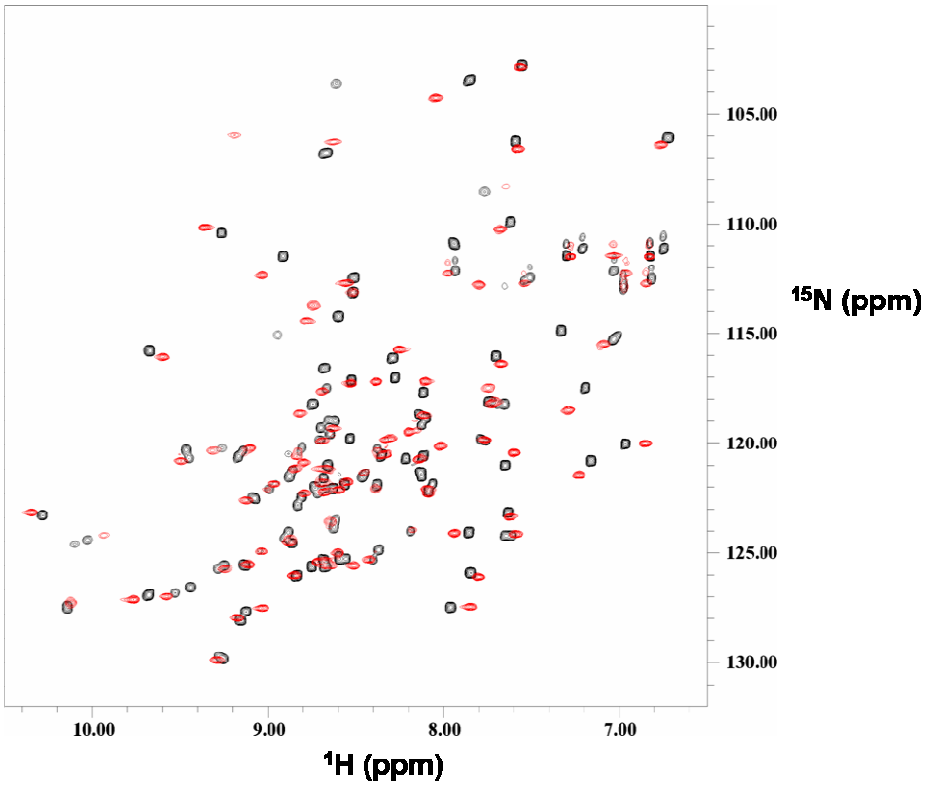

Figure 2.18. NMR spectroscopy. Overlay of the plot of the HSQC spectra of the free PDZ1 (black) and the P3 peptide-saturated form of the protein (red).

PDZ1 triggers aggregation of DLL4_IC

One important question that arises from the present findings is: does the globally disordered DLL4_IC protein acquire structure upon interaction with its PDZ target? We addressed this by titrating the ^{15}N labeled DLL4_IC protein with increasing amounts of unlabeled PDZ1, and then we recorded the HSQC spectra of the mixture. In fact, a change in the pattern of the N-H cross-peaks in the spectra in the presence of PDZ1 would point to a structural transition in DLL4_IC. Surprisingly, upon addition of substoichiometric amounts of PDZ1, the resonances of the ^{15}N-coupled protons disappear from the spectrum, with a complete disappearance observed already at a molar ratio of 1:20 (**Figure 2.19a**). It must be noticed that no precipitate could be seen in the sample after PDZ1 addition.

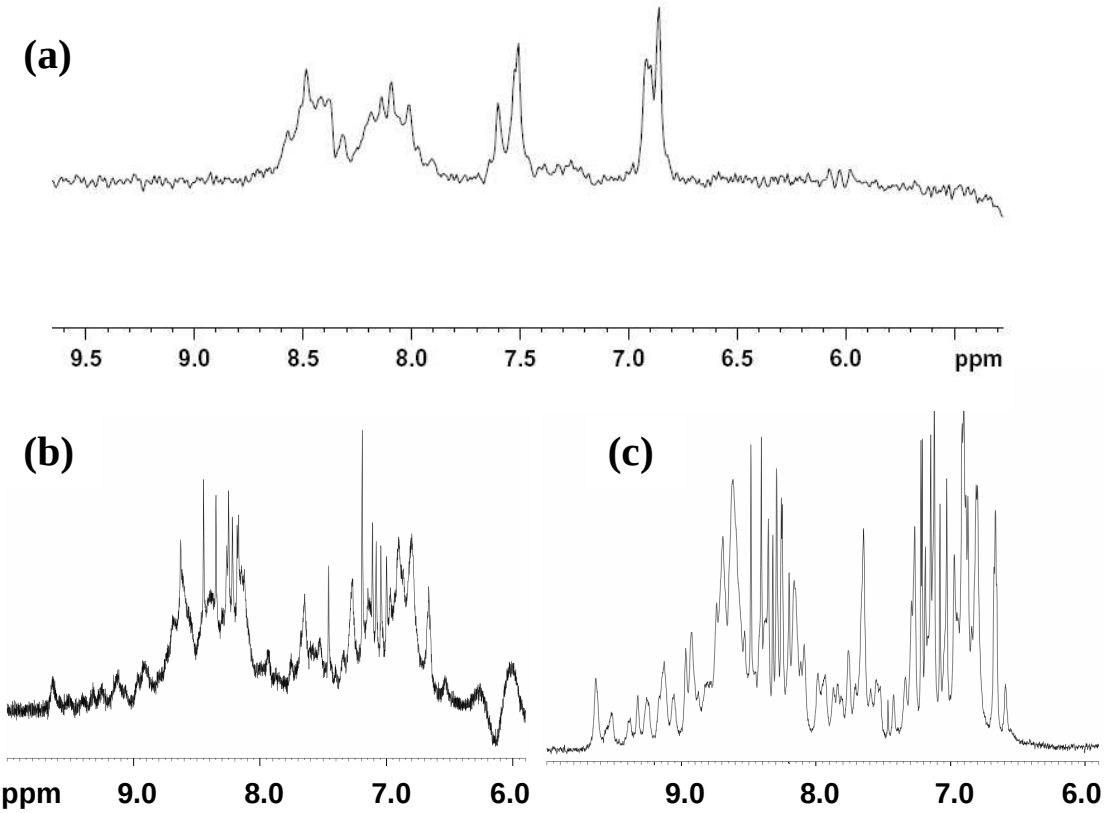

Figure 2.19.. NMR spectroscopy. (**a**) First 1D-^{1}H spectrum from the ^{1}H-^{15}N HSQC spectrum of ^{15}N-DLL4_IC (1.2 mM); (**b**) 1D-^{1}H spectrum of the DLL4_IC/PDZ1 (1:1) mixture (0.4 mM); (**c**) 1D-$^{1}H1$ spectrum of PDZ1 (0.6 mM) alone. All samples were prepared in the same buffer (20 mM phosphate, 5 mM TCEP, 100 mM KCl, pH 5.6). The spectra show the amide proton region only.

The line broadening beyond detection of the DLL4_IC N-H resonance signals is probably due to the formation of high molecular weight species that involve DLL4_IC. On the other hand, the 1D-^{1}H spectrum of the PDZ1-DLL4 mixture is very similar to that of the PDZ1 alone (**Figure 2.19c** and **d**), with the resonances in the amide region well dispersed and clearly visible: this suggests that the high molecular weight aggregates are likely to originate from oligomerization of DLL4_IC. Based on this experiment only, we cannot describe in detail how the conformation of DLL4_IC is affected by the interaction with PDZ1. Nonetheless, we can propose that, at least at the concentrations required by this experiment, the PDZ1 domain acts as nucleation site for the aggregation of DLL4_IC.

The titration of PDZ1 with the P3 peptide points to a β-strand enrichment in the peptide

Our findings suggest that PDZ1 forces DLL4_IC into an aggregated, still soluble state. In order to be able to explain the mechanism that underlies this surprising phenomenon, one should first identify which region of DLL4_IC mediates the aggregation process triggered by the PDZ1 domain. The sequence analysis and conformational studies on DLL4_IC showed that its C-terminal region (P3) has a relevant tendency to adopt a β-strand conformation that could originate from both intra-molecular (β-hairpin conformation) and inter-molecular (fibril-like conformation) interactions. A β-strand enrichment is often observed in disordered peptides that undergo aggregation processes and fibril formation [77]. Therefore, we wondered if the P3 peptide might adopt the same β-strand conformation in the presence of PDZ1: this would be consistent with the observed aggregation of DLL4_IC resulting from the interaction with PDZ1, and would suggest that the C-terminal region of DLL4_IC is the region that mediates the aggregation process.

In order to assess this, we titrated the recombinant PDZ1 domain with increasing amounts of P3 peptide, and evaluated the secondary structure content of the mixture by far-UV CD spectroscopy. The conformation of PDZ1 alone is typical of ordered proteins and the

secondary structural content is in agreement with that reported in the literature, while the P3 peptide displays a spectrum typical of disordered proteins (**Figure 2.20a**). After mixing PDZ1 with a molar excess of P3, the observed CD spectra of the mixture differ from the corresponding theoretical sum curve calculated from the individual spectra (**Figure 2.20b**). Since the theoretical curve corresponds to the spectrum that would be observed if no structural variations occur, deviations from these curves are indicative of structural transitions. In particular, an increase of the negative band at 219 nm and of the positive band at 190 nm is observed upon complex formation: these variations point to a gain in beta structure. However, the deconvolution of the spectrum of the complex and that of the theoretical sum could not discriminate any significant difference in the secondary structure content. Since these structural variations are of small entity, in the deconvolution they are probably masked by the high beta structure content of PDZ1.

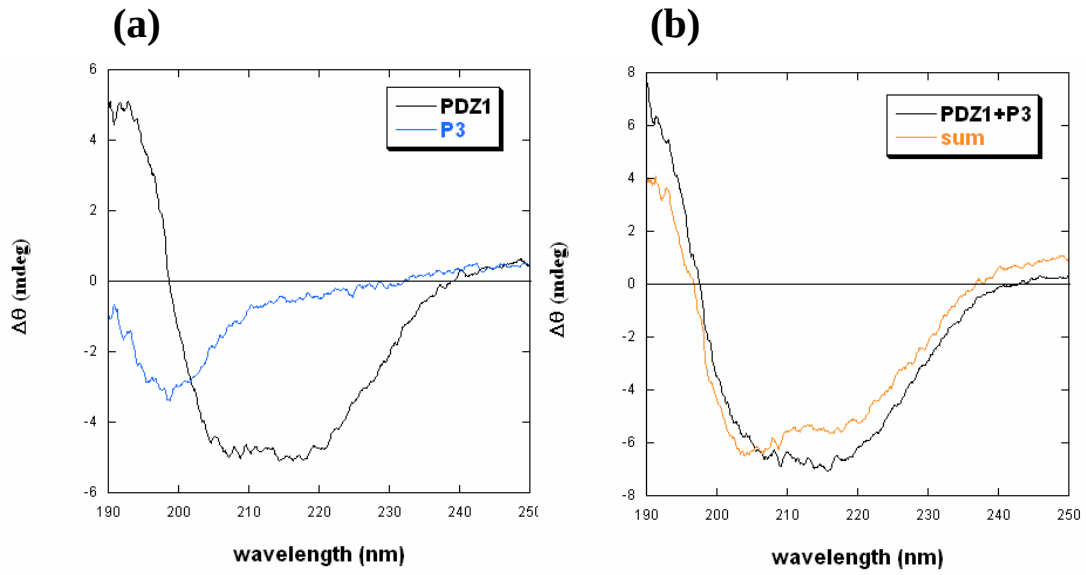

Figure 2.20. far-UV CD spectra of P3 (15 µM) and PDZ1 (10 µM) in 5 mM Tris-HCl buffer, 1 mM TCEP, alone (**a**, blue and black trace, respectively) or mixed (**b**, black trace). The theoretical spectrum of the sum of the individual spectra shown in (**a**) is the yellow trace in (**b**).

Interestingly, when PDZ1 is titrated with increasing excesses of P3, the band at 190 nm and that at 219 nm increase proportionally to the amount of peptide (**Figure 2.21a**). This

observation is surprising since, bearing in mind that the stoichiometry of the reaction is 1:1 and that the peptide alone is in the random coil conformation (**Figure 2.21b**), in these saturating conditions one would expect to see the opposite trend, with a major contribution of the negative band at 198 nm due to the unbound peptide that would decrease the positive band at 190 nm. If we assume that the P3 peptide is the flexible, plastic species that changes its conformation in the mixture with PDZ1, the spectrum obtained from the subtraction of the spectrum of the PDZ1 alone from that of the complex should represent the contribution of the peptide only (**Figure 2.21c**).

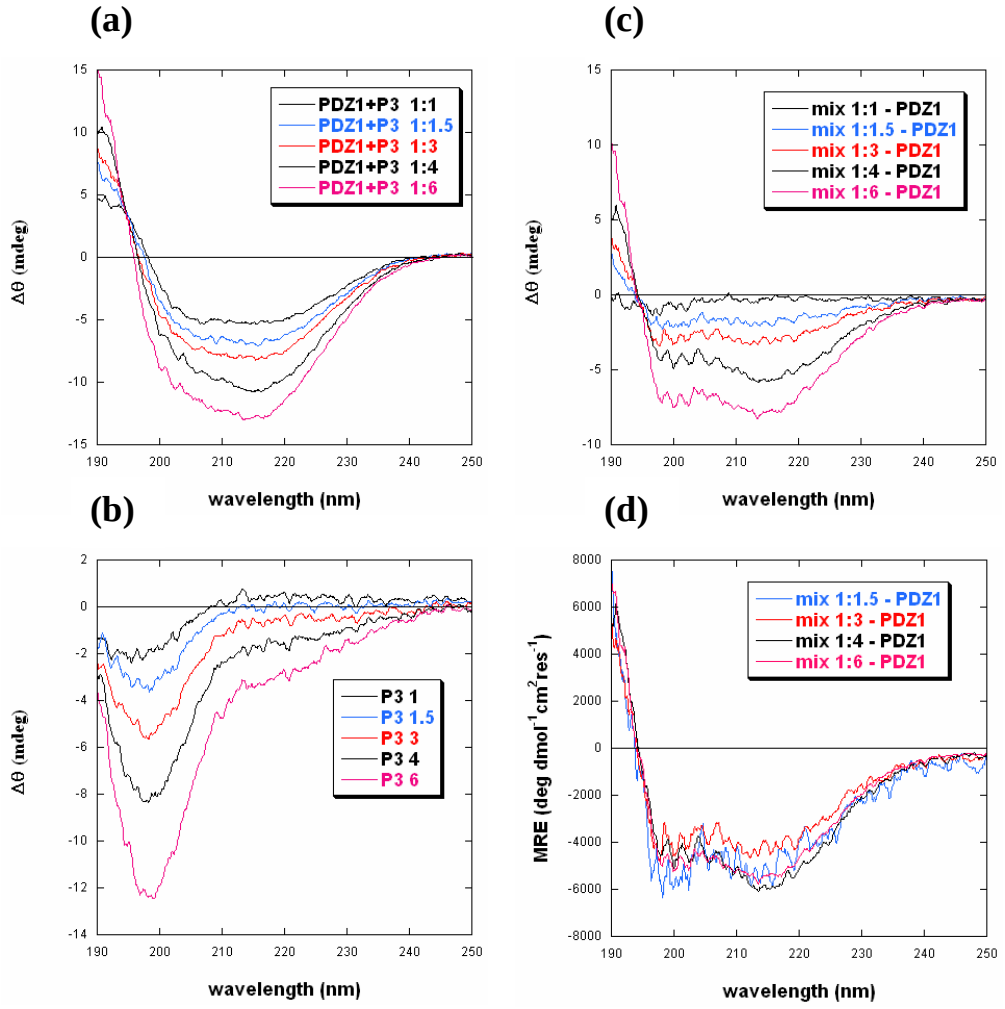

Figure 2.21. (**a**) far-UV CD spectra of PDZ1 (10 µM) in 5 mM Tris-HCl buffer, 1 mM TCEP, after the addition of P3 at the indicated molar excesses; (**b**) spectra of P3 alone at the indicated molar excesses relatively to 10 µM PDZ1, in the same buffer as in (**a**); (**c**) spectra shown in (**a**) after subtraction of that of PDZ1; (**d**) spectra shown in (**c**) after normalization.

The spectra obtained from this subtraction are significantly different from those of the peptide alone at the same concentrations (**Figure 2.21b**) and, after normalization for the peptide concentration, they all overlay (**Figure 2.21d**). These normalized spectra closely resemble those of P3 in the β-strand-like conformation observed at slightly acidic pH or in 20% TFE (**Figure 2.10f**). Since this β-strand enrichment can be induced in the P3 peptide by the presence of substoichiometric concentrations of PDZ1, we propose that the region of DLL4_IC spanned by P3, which contains the PDZ-interacting motif, mediates the aggregation observed in DLL4_IC (**Figure 2.19**) upon interaction with PDZ1. The possible implications of these findings will be discussed in the Discussion section.

2.3. Widening the view: intrinsic disorder in single-pass transmembrane receptors

The structural data on the cytoplasmic tail of DLL4 presented here show its intrinsic, globally disordered state in solution. Similarly, our previous experimental data on the intracellular region of Jagged-1 [54,55], together with the analysis of the sequences of the other Notch ligands [28], point to a disordered nature of the cytoplasmic tails of these proteins. All Notch ligands belong to the class of membrane proteins with a single transmembrane helix (single-pass) and the N-terminus located in the extracellular space (type I). In order to assess if the high incidence of protein disorder predicted and observed in the cytoplasmic region of Notch ligands is a more general phenomenon affecting other proteins of the same transmembrane class, we carried out a comprehensive computational study on the incidence and location of protein intrinsic disorder in 369 human single-pass type I transmembrane receptors.

Disorder predictions

A dataset of 369 sequences that included all human single-pass receptors was generated and subsequently divided into two subsets: the first consisted of the receptors' intracellular regions, while the second included the extracellular regions. The location of the transmembrane helices for identifying the sequences' boundaries within the subsets was assigned according to the Swiss-Prot database. DisProt [78], the database of protein disorder (469 sequences), and a reduced set (1357 sequences) extracted from SCOP [79], a database of domains of known structure, were used as control datasets. Intrinsically disordered regions in each of these datasets were predicted by subjecting the sequences to DisEMBL [80]. **Figure 2.22** shows the fraction of sequences in the datasets versus the fraction ($\geq$ 20%) of residues predicted to be disordered. First, a surprisingly high incidence of intrinsic disorder is observed in the intracellular subset, as it contains a significantly higher fraction of disordered residues relatively to the DisProt dataset. Also, the intracellular subset contains on average a higher amount of disordered residues as compared to the extracellular subset, with a relevant fraction of mainly (>50% of

disordered residues) disordered sequences (13% and 38% according to the Remark465 and Hot Loop definitions, respectively, as reported in reference 79, whereas in the latter the incidence of intrinsic disorder is only slightly higher than that computed for the SCOP domain database.

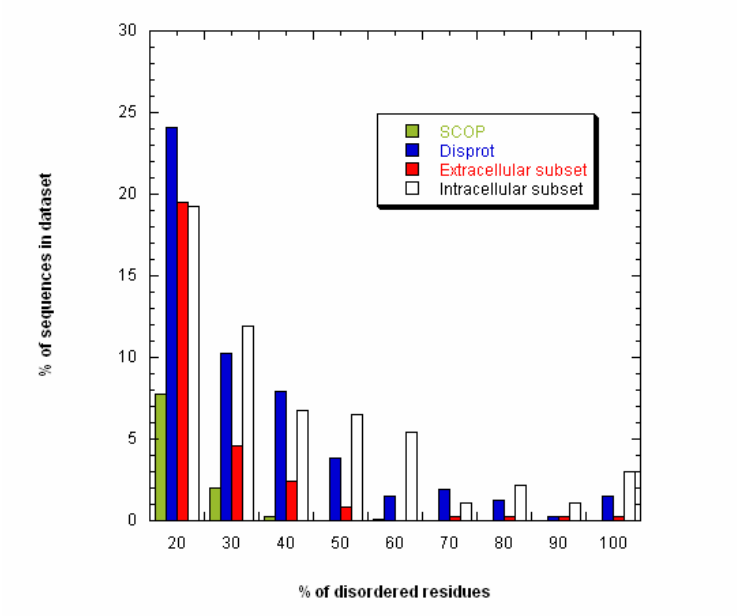

Figure 2.22. Disorder predictions. Percentage of sequences in datasets (reduced SCOP, Disprot, extracellular and intracellular regions of type I single-pass receptors) with a content of predicted disordered residues ≥ 20%. Intrinsically disordered regions were computed by DisEMBL using the Rem465 definition.

IUPred predictions are consistent with DisEMBL results. The fraction of disordered residues is higher in intracellular regions, as calculated from pairwise energies. Overall, the % of residues with a IUPred score above the threshold of 0.5 is 29 and 10 for intra- and extracellular regions, respectively. Very similar results were obtained using either the "long" or "short" disorder definition (**Figure 2.23**).

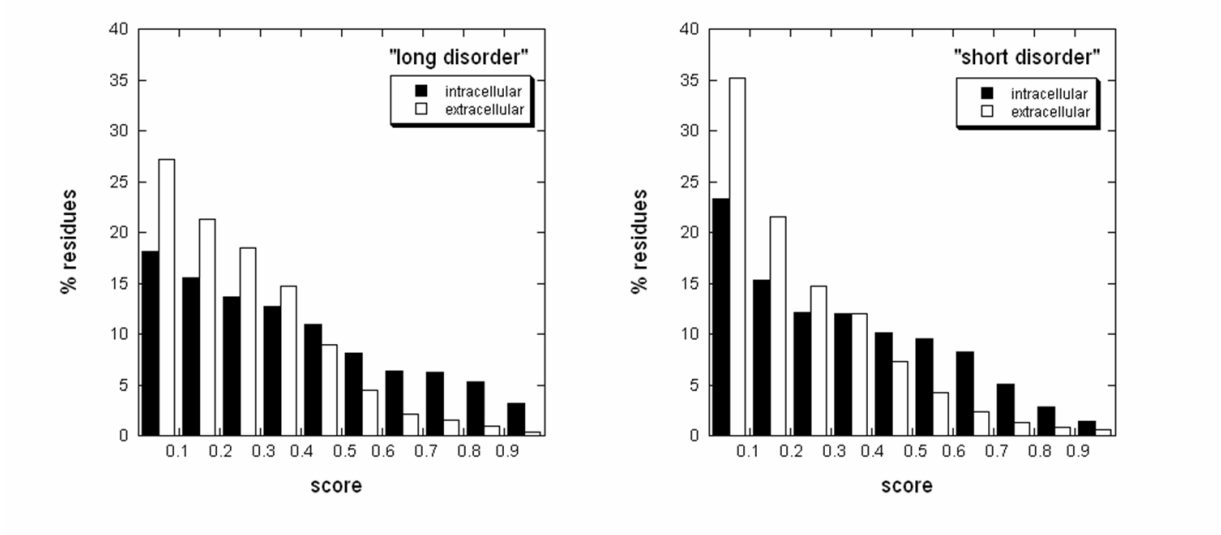

Figure 2.23. Disorder predictions. The percentage of residues in the intra- (filled bars) and extracellular (empty bars) regions of human single-pass transmembrane receptors with the type I topology is plotted versus the IUPred score, calculated using either the "long" (top) or the "short" (bottom) disorder definition.

Charge/Hydropathy plot

A combination of low mean hydropathy and relatively high net charge was shown to represent an important prerequisite for the absence of compact structure in proteins under native conditions [81]. Consequently, intrinsically disordered proteins generally localize within a unique region of the charge/hydropathy phase space, whereas ordered proteins cluster in a separate region (grey and green shaded areas in **Figure 2.24**, respectively) [82]. In **Figure 2.24**, we plotted the absolute value of the mean net charge versus the mean hydropathy values for the sequences in both the intracellular and extracellular subsets. Clearly, the mean net charge/hydropathy of the intracellular regions is broadly distributed in the phase space, as compared to that of their extracellular counterparts, with an averagely higher mean net charge and a relatively lower mean hydropathy, suggesting that the intracellular regions are more prone to intrinsic disorder. This is consistent with the abovementioned disorder predictions.

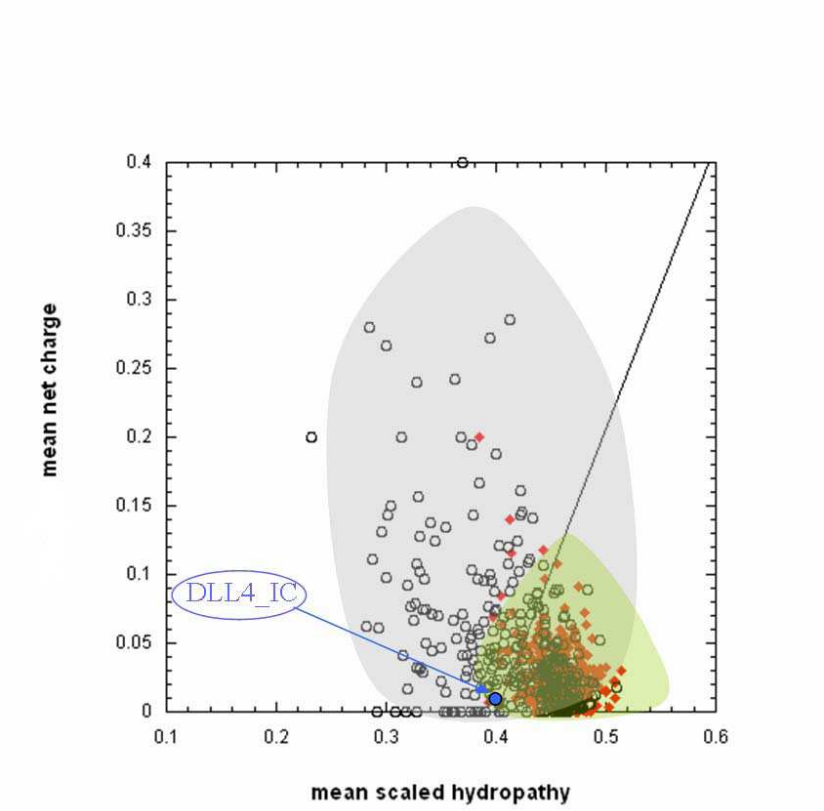

Figure 2.24. Charge/Hydropathy plot. The absolute value of the mean net charge is plotted versus the mean scaled Kyte-Doolittle hydropathy for the Intracellular Subset (circles) and Extracellular Subset (diamonds). The shaded areas contain 90% of the points from the dataset of disordered regions (DisProt, grey) and the dataset of domains of known structure (SCOP, green). The border between structured and natively disordered proteins is drawn as a line. The charge/hydropathy value relative to DLL4_IC is also shown.

Amino acid compositional analysis

The amino acid compositional analysis confirms these observations. **Figure 2.25** shows a comparison of the amino acid compositions of the intra- and extracellular subsets along with the comparison between the two control datasets (ordered and disordered proteins). With few exceptions (see Discussion), the intracellular set is depleted in the order-promoting residues [83] (W, C, F, I, Y, V, L, N) and enriched in the disorder-promoting residues (A, R, G, Q, S, P, E, K) compared to the extracellular set. The same trend is observed in the comparison between the datasets of ordered and disordered proteins.

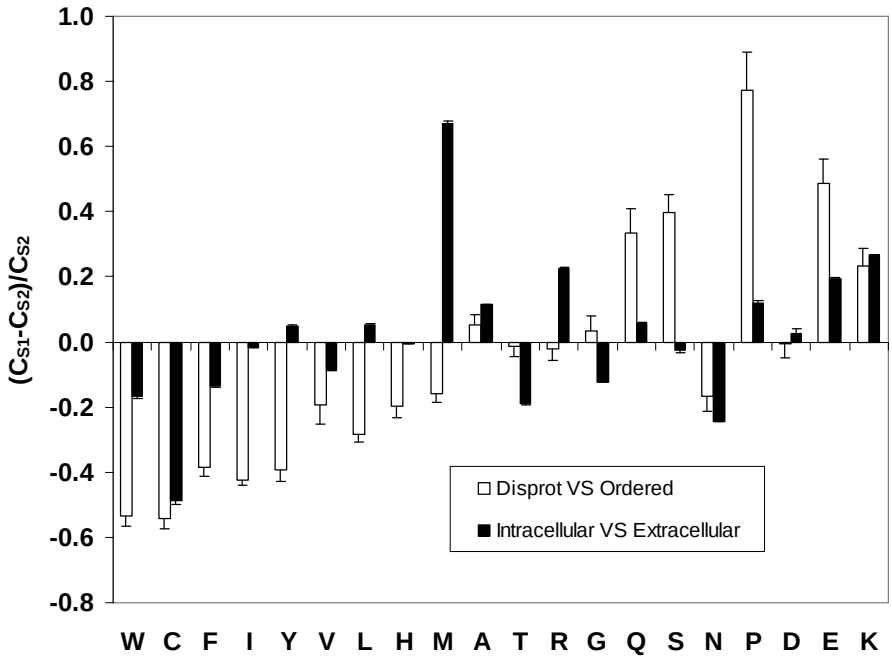

Figure 2.25. Amino acid compositional analysis. Enrichment or depletion in each amin acid type appears as a positive or negative bar, respectively. Amino acids are indicated by the single-letter code and ordered according to increasing flexibility. Order-promoting residues: W, C, F, I, Y, V, L, N; disorder-promoting residues: A, R, G, Q, S, P, E, K; undefined: H, M, T, D. Error bars are also shown.

Taken together, our disorder predictions, charge/hydropathy and compositional analyses strongly suggest a major incidence of intrinsic disorder in human single-pass transmembrane receptors, and that intrinsic disorder is highly concentrated in the cytoplasmic region.

3. Discussion

3.1. Intrinsic global disorder and inducible local order in the intracellular region of DLL4: implications in function

From a biophysical perspective, proteins fall into a structure continuum (**Figure 3.1**), going from tightly packed globular domains, to less ordered structures such as folded domains joint by flexible linkers, to collapsed states with a residual content of transient secondary structures called "molten globules" to, finally, proteins that in their native, intrinsic state appear as highly extended dynamic ensembles.

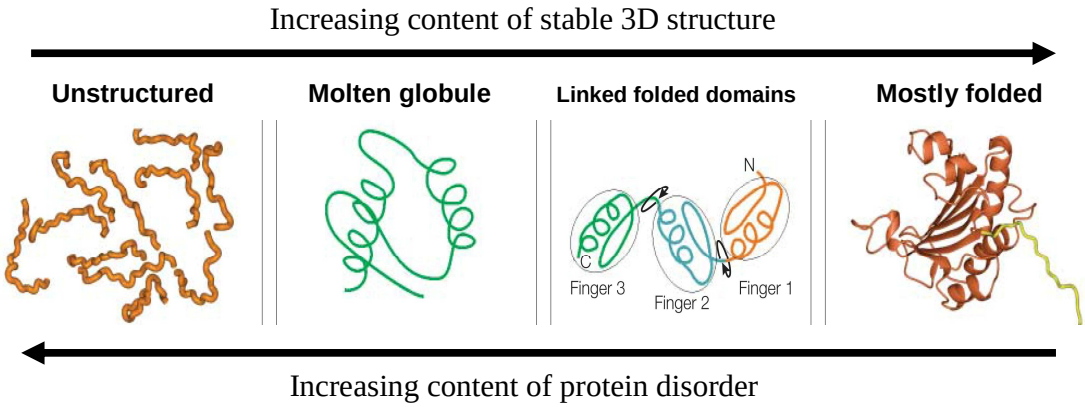

Figure 3.1. The protein structure continuum (figure adapted from reference 85).

Intrinsic disorder in proteins has been shown to be a widespread phenomenon by both computational and experimental methods and it is now recognized that a certain protein not only can be functional without having a defined three-dimensional structure, but that its functionality can lie indeed in its being disordered [84,85]. Extensive disorder predictions to sequences of genomes of increasing complexity revealed that disorder is the highest in eukaryotes [86] and that long disordered regions are mostly found in proteins involved in regulation and signaling [87]. These two observations were interpreted as the consequence of the increased complexity of the regulatory networks in multicellular organisms, where

the plasticity of disordered regions in highly connected proteins is exploited to bind different signaling partners through disorder-to-order transitions [88].

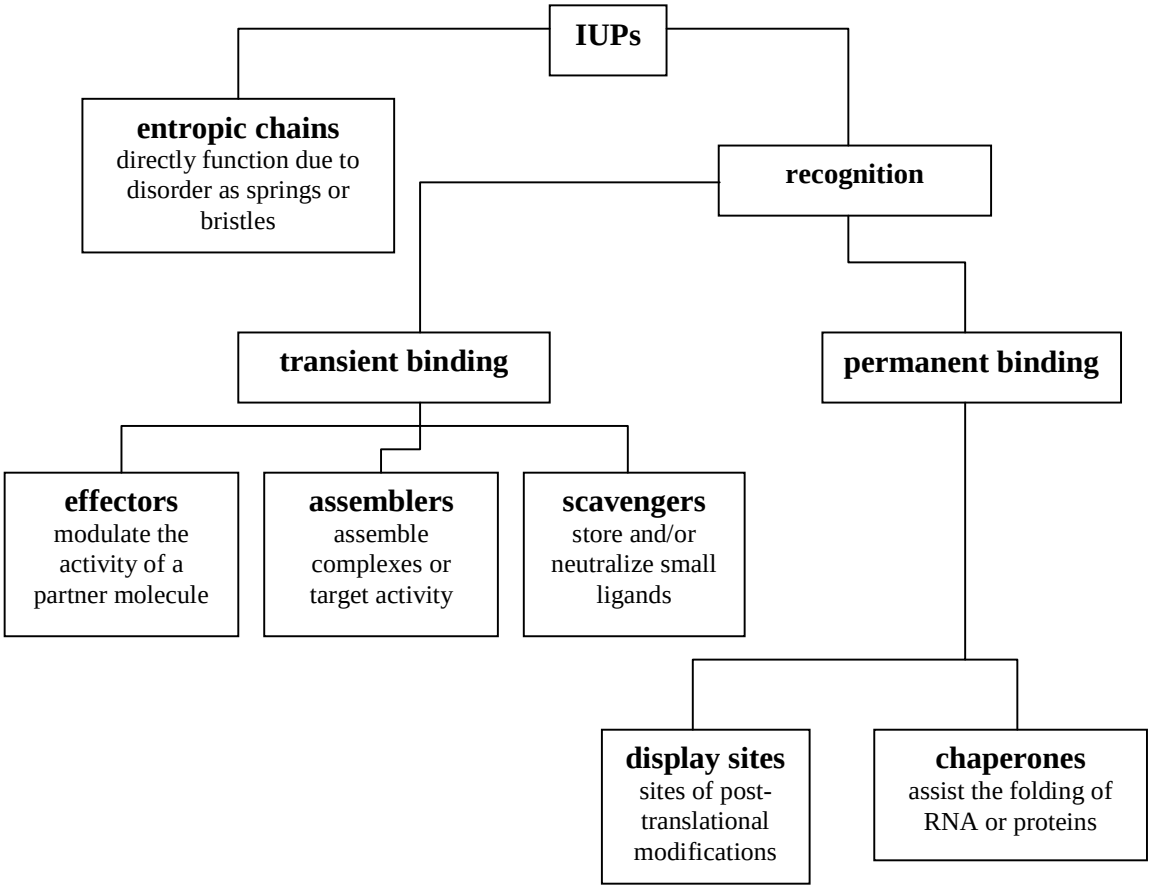

Figure 3.2. Functional classification scheme of IUPs. The function of IUPs stems either directly from their capacity to fluctuate freely in a large conformational space (entropic chain functions) or the ability to transiently or permanently bind partner molecule(s). Our data suggest that intrinsic disorder in DLL4_IC may play a role in displaying motification sites (global disorder) as well as in recognition and binding to protein partners (inducible local order). This figure was adapted from reference 89.

Intrinsically disordered proteins (IUPs) are usually characterized by a high number of charged residues compared to the number of hydrophobic residues, which results in the lack of a hydrophobic core, little or no secondary structure elements, high hydrodynamic radius, and often a high net charge at physiological pH. From the biophysical point of view, IUPs can be considered as polypeptide chains that in physiological conditions are sampling a much wider conformational space with respect to globular proteins. It has been proposed that this extended sampling can indeed have several advantages. IUPs have a much larger interaction surface/volume ratio compared to globular proteins, which allows for the accommodation of a relatively high number of docking sites on a relatively short polypeptide chain, at the same time reducing the protein volume, therefore the molecular crowding. The extended conformational sampling has interesting thermodynamic consequences. It enables IUPs to couple folding to binding maintaining high specificity and low affinity due to the balance between the enthalpic contribution to binding and the opposite entropic effect. Indeed, weak although specific interactions are most important in molecular recognition. The “folding upon binding” mechanism is especially advantageous in signaling and regulation contexts, where proteins must associate and quickly dissociate when the signaling is over.

The Notch ligand DLL4 is a transmembrane protein composed of a globular extracellular domain responsible for receptor binding and of a short cytoplasmic tail which is a structural requirement for Notch bi-directional signaling, for the cross-talk with other signaling pathways, for cell-autonomous, Notch-independent signaling, and for endocytosis-mediated receptor shedding. The cytoplasmic tail of DLL4 is evolutionary very well conserved (**Figure 3.3**) and does not share any homology with globular domains of known structure, therefore the main question we posed was: does it encode a yet unknown functional fold or is it entirely or partly unstructured? Our disorder predictions together with NMR, CD and size exclusion chromatography data on a recombinant purified protein representing the intracellular region of Notch ligand DLL4 (DLL4_IC) are consistent with a globally disordered, yet quite collapsed state of the protein in solution. Therefore, the structural behavior of this protein fragment as a whole resembles that of a native molten globule. The observation of inducible transient local

structures in DLL4_IC reinforces this view. In fact, despite its lack of a stable tertiary structure, our sequence and structural analyses reveal that DLL4_IC has the inherent propensity to form local secondary structures. The structural mapping using synthetic peptides that map to different regions of the DLL4_IC sequence showed that the secondary structures mainly form in the plastic C-terminus of the protein, that they are reversible and inter-convertible and can be specifically induced, from a disordered conformation, through well defined changes in the physico-chemical environment.

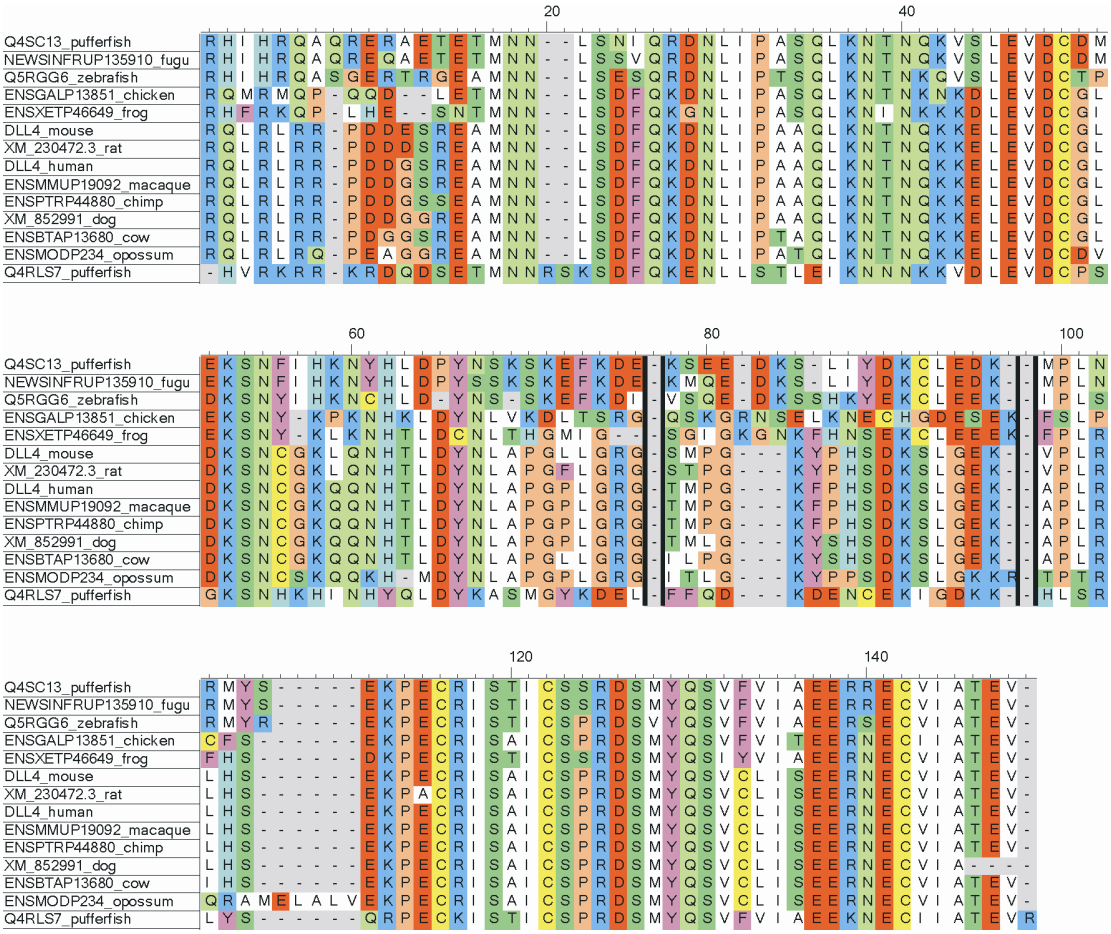

Figure 3.3. Sequence alignments. The sequences of the intracellular tail of the indicated DLL4 hortologues. were aligned using ClustalW and colored using CINEMA. Acidic residues (D, E) in red; basic (K, R) in blue; histidines (H) in light blue; aliphatic (A, V, L, I, M) in white; small hydrophobic (G, P) in orange; aromatic (F, Y, W) in magenta; hydroxyl-containing (S, T) in dark green; amide containing (N, Q) in light green; cysteines (C) in yellow. This figure was taken from reference 28.

The coil-to-strand transition

The secondary structure predictions all point to the presence of a short stretch of β-strand structure in the C-terminus of DLL4, involving the last twenty amino acids (**Figure 3.4**). This region is very well conserved through species and is of functional relevance, as it includes the C-terminal tetrapeptide (ATEV) shown to be required for the interaction with the DLL4 physiological partner Dlg-1.

```
                          CCCCCEEEEEEEECCCCCEEEEECC

DLL4_human                RDSMYQSVCLISEERNECVIATEV-
ENSMMUP19092_macaque      RDSMYQSVCLISEERNECVIATEV-
ENSPTRP44880_chimp        RDSMYQSVCLISEERNECVIATEV-
XM_852991_dog             RDSMYQSVCLISEERNECVIA----
DLL4_mouse                RDSMYQSVCLISEERNECVIATEV-
XM_230472.3_rat           RDSMYQSVCLISEERNECVIATEV-
ENSBTAP13680_cow          RDSMYQSVCLISEERNECVIATEV-
ENSMODP234_opossum        RDSMYQSVCLISEERNECVIATEV-
ENSGALP13851_chicken      RDSMYQSVFVITEERNECIIATEV-
ENSXETP46649_frog         RDSMYQSIYVIAEERNECVIATEV-
Q5RGG6_zebrafish          RDSVYQSVFVIAEERSECVIATEV-
Q4SC13_pufferfish         RDSMYQSVFVIAEERRECVIATEV-
NEWSINFRUP135910_fugu     RDSMYQSVFVIAEERRECVIATEVR
```

Figure 3.4. Secondary structure predictions (C, coil; E, strand; H, helix) and sequence alignment of the C-terminal region of DLL4 homologues.

The predicted β-strand conformation is in fact observed by far-UV CD spectroscopy, in a peptide representing the last 24 amino acid of DLL4 (P3) in water at slightly acidic pH or in the presence of 20% of the co-solvent TFE. In addition, our data showed that a similar β-strand conformation can be acquired by the peptide in the presence of substoichiometric amounts of the first PDZ domain of Dlg-1, which we proved to be the specific determinant of Dlg-1 that binds DLL4_IC *in vitro*. This phenomenon can be interpreted in different ways, with accordingly different thermodynamic implications important for the binding of DLL4 to Dlg-1. The regulation of the DLL4/Dlg-1 interaction is functionally relevant, as its proposed role is to recruit Dlg-1 at the cell membrane, tightening cell contacts and reducing cell motility [30].

(i) the β-strand conformation is intra-molecular (β-hairpin) and preformed in the intracellular environment

In the available crystal structures, peptides bearing a PDZ recognition motif bind in a groove between βB and αB on the surface of the PDZ domain and adopt a β-strand conformation, stitched as an additional strand into the antiparallel β-sheet on the surface of the PDZ domain [72]. Therefore, a stable β-hairpin at the C-terminus of DLL4, or a significant population of molecules in the same conformation, would present the ATEV motif in the correct geometry to dock its PDZ target, drastically reducing the entropic cost of binding. This view is supported by the CD analysis of the titration of PDZ1 with the P3 peptide, in which we show that the fully formed β-strand conformation is observed in P3 even at substoichiometric concentrations of PDZ1.

Alternatively, the β-hairpin could be induced upon binding with the PDZ domain, generating a complex with a lower affinity. Subtle changes in the concentration of DLL4 would direct the binding to Dlg-1 towards one mechanism or the other, thereby tuning the formation of the complex.

Whether a β-hairpin in the C-terminus of DLL4 is preformed in solution could be elucidated by recording and analyzing the 2D TOCSY and NOESY NMR spectra of the P3 peptide at pH~6 or in the presence of deuterated TFE. It would then be possible to (a) investigate the identity, extent, and location of secondary structural elements in the peptide quantitatively, based on the inspection of the ^{13}Cα and ^{1}Hα resonance assignments (chemical shift index determination), (b) identify the inter-strand Hα-Hα and Hα-NH interactions through the analysis of the NOEs, (c) identify the hydrogen-bonded amide hydrogens by plotting their chemical shift variations over temperature changes (temperature coefficient determination) and (d) assign the peptide residues to a particular secondary structure element by measuring the $^{3}J_{NH\text{-}HA}$ coupling constants.

Conversely, ^{15}N and ^{13}C edited experiments on the complex between a doubly labeled PDZ protein and the P3 peptide would be needed to clarify if the β-structure in the peptide is induced in the bound form. Pulse sequences which can filter out the ^{15}N and ^{13}Cα-bound hydrogen resonances of the PDZ would allow the observation of the peptide resonances only. On the other hand, the analysis of the NOEs from the labeled species only, would make possible to assign the residues of the peptide that actually make contact

with the PDZ domain, and establish if they take part in the formation of some secondary structure.

Also, after assignment of the PDZ resonances, the chemical shift tracking of the PDZ ^{15}N-^{1}H cross peaks upon addition of P3 peptide would allow the determination of the surface of the PDZ that interacts with the peptide. This would tell if the interaction involves the canonical binding groove only, as observed with other PDZ-interacting shorter peptides, or if it extents beyond to other regions of the PDZ domain. This information would be potentially important, as additional contacts to PDZ residues away from the binding groove are expected to increase the specificity of the binding.

(ii) the β-strand conformation is inter-molecular (fibril-like) and triggered by the PDZ protein

This scenario is supported by the NMR data that show that substoichiometric amounts of PDZ1 protein are sufficient to broaden the DLL4_IC N-H signals beyond detection, suggesting the formation of high molecular weight oligomers of DLL4_IC molecules. According to this view, the exposed β-strand in the binding groove (βB) of the PDZ domain would function as a template that constrains the C-terminus of DLL4 into a β-strand conformation. This constrained DLL4_IC molecule would in turn propagate the formation of the same β-strand structure to other DLL4_IC molecules, generating high-molecular weight species. It is tempting to speculate that this phenomenon may be somehow connected to the observed clustering of the DLL4/Dlg-1 complex at cell-cell junctions [30]. However, this could be also a non-specific phenomenon that originates from the relatively high protein concentrations required by NMR. Other experiments are needed to test this hypothesis. For instance, running the DLL4_IC/PDZ1 mixture at different molar ratios through a calibrated size exclusion chromatography column would elucidate if species with molecular weights higher than expected are actually formed.

The coil-to-helix transition

The other major local conformation observed in DLL4_IC is helical and can be induced with SDS micelles in a slightly acidic buffer. Our data show that DLL4_IC weakly associates with micelles while gaining helical structure and are consistent with the formation of a ~10 residue α-helix at saturation. Although all secondary structure predictions agree with a high propensity of the N-terminus of DLL4_IC to be helical, our results prove instead that the pH-dependent helix formation observed with SDS micelles involves the 24 carboxy-terminal amino acids (P3 peptide). We believe that these conformational changes are coupled with the protonation of one or more acidic residues in the C-terminal region of DLL4_IC. It is not likely, however, that this partial charge neutralization is sufficient to promote complete insertion of the peptide in the hydrophobic layer, given the presence of hydrophilic and positively charged residues in the sequence spanned by the P3 peptide. SDS micelles provide a hydrophobic/hydrophilic interface and can be considered a first approximation of biological lipid membranes. Therefore, the mainly helical conformation of DLL4_IC in the presence of SDS may be representative of the membrane-bound, uncomplexed form of DLL4. The possible biological implications of the pH-dependent conformational change are not known yet. Whereas different cell compartments can be associated with different pH values, little is known of the biophysical properties of the membrane-cytoplasm interface [90]. In an early study, fluorescein was used to map the pH distribution in yeast cells, and it was proposed that the intracellular pH is not homogenous, but decreases to ~6.0 in proximity of the membrane [91]. The pH gradient between the membrane interface and the cytosol would be generated by the negatively charged head groups of phospholipids present in the membrane of eukaryotic cells. The partial folding of the C-terminus of the cytoplasmic tail of DLL4, accompanied by its association with the inner side of the cell membrane, may have relevant effects on the function of the ligand in Notch signaling. For instance, it would selectively mask certain residues that are potential targets for post-translational modifications such as phosphorylation, ubiquitination and glycosylation, and at the same time it would leave others exposed for the same modifications. In a similar way, it would mask or expose selected binding motifs with respect to binding partners. For instance, it

could prevent the interaction with Dlg-1 by masking the ATEV motif. Further studies are needed to confirm this hypothesis. In particular, synthetic membranes which can reproduce the composition and geometry of biological membranes with a higher accuracy (*e.g.* liposomes composed of mixtures of biomembrane phospholipids), should be used in the above studies.

Alternatively, the secondary structures observed in the C-terminus of DLL4 might be induced through the binding of DLL4's cytoplasmic tail to other still unknown cytoplasmic or nuclear partners, before or after it is cleaved and released as a signaling fragment. In this context, the intracellular domain of the DLL4 homologue DLL1 was recently shown to act as a transcription-cofactor in the signal-sending cell, as it mediates TGF-β/Activin signaling through binding to Smad proteins in the nucleus [48].

The structural randomness present in intrinsically disordered regions is intuitively associated with a lack of evolutionary conservation in the amino acid sequence. In fact, because disordered regions do not require structural constraints for folding, they have a higher degree of mutability as compared to folded domains [92]. Expansion of genetically unstable repeats has been proposed as a major mechanism for generating novel genetic material encoding long disordered regions [93]. Bearing in mind this general rule, it seems somehow paradoxical that a globally disordered functional region such as DLL4_IC is on the other hand very well conserved across species. **Figure 3.3** shows how the conservation extends well beyond the PDZ interacting C-terminal motif. We believe that this apparently contradicting observation is just the evidence of the intrinsic functional role of structural disorder in DLL4_IC, which seems to have been "selected" in evolution, through the conservation of the DLL4_IC amino acid sequence. In other words, while sequence conservation in DLL4_IC suggests that precise consensus motifs are required for specific patterns of post-translational modifications to take place and for specific protein-protein interactions to occur, the conservation of structural disorder indicates disorder as a fundamental prerequisite for these events to occur. According to this view, we propose that the co-existence of global disorder (i.e., lack of a well-defined globular structure) and local pre-organization (i.e., the propensity to form certain types of

secondary structures locally, either in a stable or transient way), may indeed represent the mechanism exploited by DLL4 to carry out its action in both the Notch-dependent and independent processes it mediates. Global disorder can impart the structural flexibility required to expose specific sites for posttranslational modifications (phosphorylation, ubiquitination or glycosylation) while local pre-organization can guarantee specificity to the binding of DLL4 to its intracellular partners, such as Dlg-1. Further studies aimed at determining the actual post-translational modifications that DLL4 undergoes, the entire spectrum of its interactors together with a structural characterization of the interactions, are needed to corroborate this view.

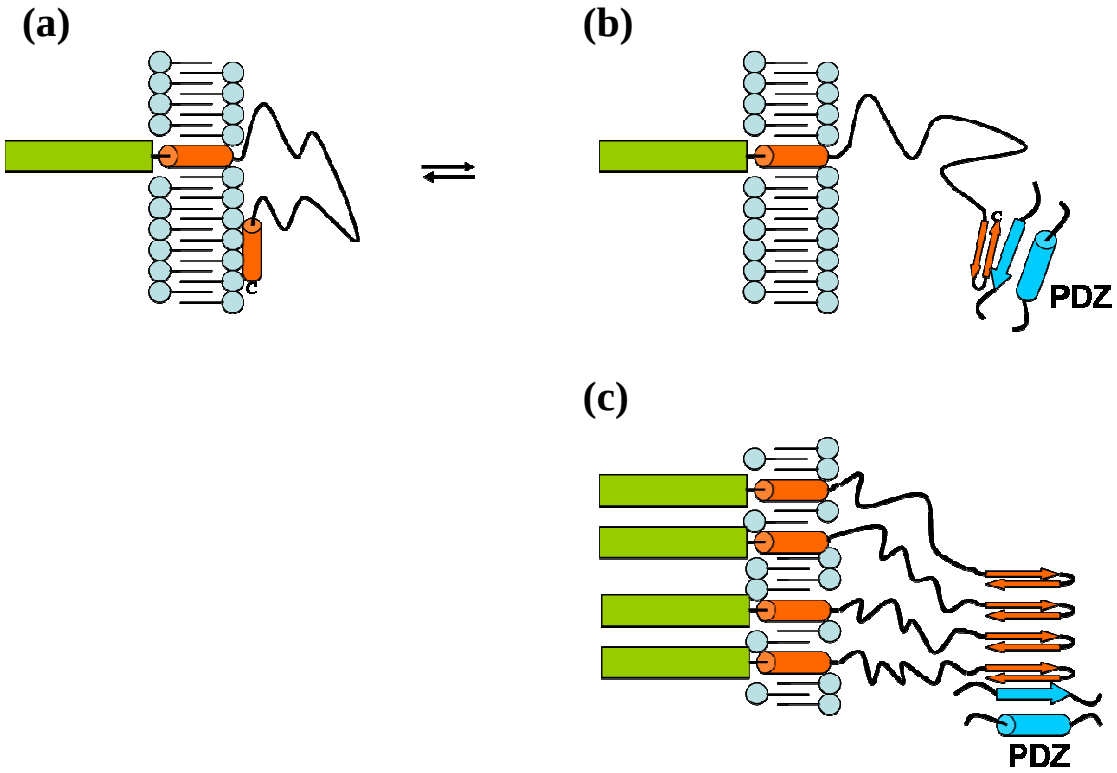

Figure 3.5. Cartoon representation of the possible conformations of the intracellular region of Delta-4 at the membrane/cytoplasmic interface; (**a**) the intracellular tail of Delta-4 is disordered except its extreme C-terminus which is helical when bound to the inner side of the lipid bilayer; (**b**) the β-strand structure in the C-terminus is intra-molecular and either preformed or induced by the presence of its PDZ target; (**c**) the β-strand conformation is induced by the PDZ domain and propagates inter-molecularly causing the oligomerization of Delta-4; α-helices and β-strands are represented by cylinders and arrows, respectively; the PDZ binding groove is colored in blue.

3.2. Protein intrinsic disorder in the cytoplasmic tail of single-pass transmembrane proteins: a conserved functional role?

Single-pass transmembrane receptors play an important role in cell communication and signal transduction. In the simplest functional model, upon binding of a signaling molecule to the extracellular region of a receptor, a response is initiated on the inner side of the membrane. This response is in fact accomplished by the receptor's intracellular tail. An analysis of the modular domain architecture of the human single-pass transmembrane receptors dataset (369) by the SMART tool [4] revealed that the majority of the ligand-binding extracellular regions are composed of known globular domains. On the other hand, SMART failed to identify any known domains in 63% of the intracellular regions, with the notable exception of receptors containing a kinase domain in their cytoplasmic region. This result can be due to either the presence of yet unidentified globular domains, or the prevalence of disorder in the intracellular tails of receptors. Our findings indicate that ID in the human single-pass transmembrane receptors with the type I topology is indeed predominant within the cytoplasmic regions, whereas the extracellular regions behave more like ordered proteins. These results are supported by the DisEMBL and IUPred predictions, by the plots of mean net charge versus hydropathy, and by the amino acid compositional analysis. When considering the mean net charge versus hydropathy plot, it should be kept in mind that kinase domains appear with a certain frequency in the intracellular region of single-pass receptors, which justifies the occurrence of several entries in the right-hand part of the plot, corresponding to ordered proteins. Incidentally, we remarked that there is a significant difference in the mean net charge of the extracellular and intracellular regions of the human transmembrane proteins analyzed in this work. While the majority (81%) of the extracellular regions have a negative mean net charge, the intracellular regions are nearly equally distributed between positively (47%) and negatively (53%) charged. In fact, most type I transmembrane proteins bear a short stretch of positively charged residues in their cytoplasmic tail, in the region close to the inner side of the membrane and protruding from it. As the inner leaflet of the membrane in eukaryotic cells is negatively charged, the presence of a positively charged segment is supposed to be a signal that drives the protein into the correct orientation [94].

With respect to the amino acid composition, the interpretation of the results is more complex. While the intracellular regions are depleted of several order-promoting residue types, such as W, C, F, and V, and enriched in some of the disorder-promoting ones such as E and K, the trend observed in the DisProt versus Ordered comparison is not always respected. For example, methionine, which belongs to the order-promoting residues, displays a high frequency in intracellular regions. It is likely that methionine, as it is susceptible to oxidation, prefers the reducing environment of the cytosol, compared to the extracellular environment. The different redox potential in the cytoplasm and in the extracellular space may also be associated with the different distribution of other amino acid types, like cysteine, which is most frequently found in its oxidized half-cystine, structure stabilizing form in the extracellular space, and almost exclusively found in its reduced form, often coordinated to metal ions, in the intracellular space. Tyrosine, which is also an order-promoting amino acid type, is over-represented rather than depleted in intracellular regions. We think that a possible explanation is given by the fact that Y is a target for phosphorylation, known to be one of the main signaling mechanisms. The amino acids E, Q, P, and S are largely over-represented in disordered regions, but not quite so in the intracellular regions of receptors. This is due to the fact that these residues are often found in low complexity tracts (poly-glutamic acid, poly-glutamine, poly-proline and domain linkers, respectively). The intracellular regions appear to be disordered, but not compositionally biased in this sense.

Several MIRR (Multichain immune recognition receptors) cytoplasmic domains belong to the subset of sequences which predictions failed to ascribe to domains of known structure, and they were experimentally proven to be intrinsically disordered [95]. Previously, we showed that the cytoplasmic tails of all five human Notch ligands are very well conserved within ligand types, display little sequence similarity between one another, display no homology with sequences of known fold and are predicted to be disordered [28]. Moreover, we have shown that a recombinant protein corresponding to the intracellular region of the Notch ligand Jagged-1 is actually disordered in solution, while it partially folds upon interaction with synthetic membranes made of negatively charged

phospholipids [54,55]. Jagged-1, similarly to DLL4, couples Notch signaling to PDZ bearing proteins. However, differently from what observed for DLL4, the interaction with its target PDZ domain does not seem to affect Jagged-1`s globally disordered conformation (unpublished data). The secondary structure predictions carried out on Jagged-1 sequence differ from those obtained for DLL4 and the other ligands [25]. Significantly, predictions of posttranslational modification and protein-protein interaction sites are different for each of the five ligands (**Figure 3.6**). While global disorder seems to be a common feature shared by all ligands, we believe that the presence of specific patterns of pre-organized secondary structures as well as of post-translational modification and protein-protein interaction motifs might underlie the functional diversity actually displayed by the ligands.

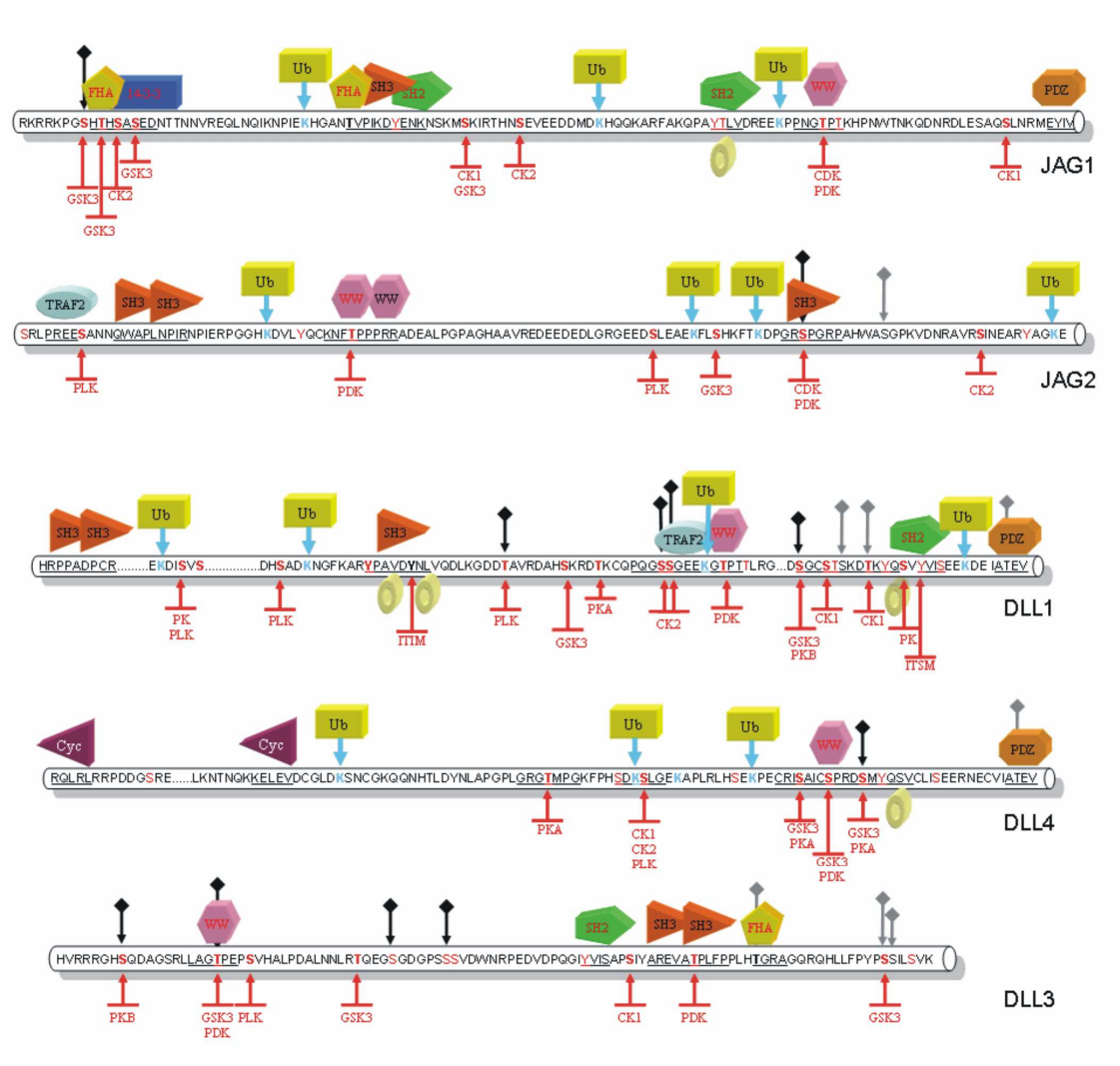

Figure 3.6. Functional analysis. Potential binding sites and post-translational modifications predicted by ELM [96,97], NetPhos [94], and O-glycosylation [98] for the cytoplasmic tail of human Notch ligands. Prediction of ubiquitination sites is based on the preference for acidic residues adjacent to the target lysine [99]. 14-3-3, 14-3-3 proteins interacting motif (Ser/Thr phosphorylation required); Cyc, cyclin binding site; FHA, forkhead-associated domain interaction motif 1 (Thr phosphorylation required); PDZ, class I, II, or III PDZ binding motif; SH2, Src Homology 2 (SH2) domains interaction motif (tyrosine phosphorylation required; subtypes include GRB2, SH-PTP2, SRC, STAT3, STAT5, STAT6); SH3, SH3 domains binding motif (subtypes include class I, class II, and other non-canonical motifs); TRAF2, tumor necrosis factor receptor associated protein binding motif; Ub, ubiquitination site; WW, WW domain binding motif (subtypes include Group I (PPXY), Group II (PPLP), Group III, and Group IV, which requires Ser/Thr phosphorylation). Tyrosine-based sorting signals responsible for the interaction with the μ subunit of the AP (Adaptor Protein) complex are shown as doughnuts. Potential phosphorylation sites are in red; kinases are abbreviated as follows: CDK, Ser/Thr cyclin dependent kinase; CK1, casein kinase 1; CK2, casein kinase 2; GSK3, glycogen synthase kinase 3; PKA, protein kinase A; PKB, protein kinase B; PDK, Proline-Directed Kinase; PLK, Polo-like-kinase. ITIM, immunoreceptor tyrosine-based inhibitory motif (tyrosine phosphorylation required); ITSM, immunoreceptor tyrosine-based switch motif (tyrosine phosphorylation required). Sites that are candidates for O-glycosylation with β-N-acetylglucosamine are shown as grey diamonds; sites that are predicted to be both glycosylated and phosphorylated are shown as black diamonds. This figure was taken from reference 28.

Overall, there is convergent evidence that the asymmetric disorder distribution observed in DLL4 and its homologues reflects a more general phenomenon in transmembrane receptors of the same class, where the extracellular domains are ordered and appear to act as rigid scaffolds for ligand binding, while the intracellular tails, which transduce the signal within the cell and activate the complex cellular response, have a higher content of structural disorder. Consequently, we speculate that the intracellular tails carry out their function by binding possibly multiple partners, either remaining unstructured, by exploiting pre-formed local secondary structures, or through disorder-to-order transitions. Further experimental data on the structural characterization of the interactions between the cytoplasmic tails of transmembrane receptors and their partners are needed in order to explore this intriguing hypothesis.

Materials and Methods

DLL4_IC sequence analysis

The DLL4_IC protein sequence was submitted to the PONDR server (http://www.pondr.com) using the default predictor VL-XT [59], the DisEMBL server (http://dis.embl.de), and the IUpred [60] server (http://iupred.enzim.hu/). Secondary structure predictions (PSIPRED, JNet, SSpro) [56-58] were run from the PHYRE web server (http://www.sbg.bio.ic.ac.uk).

Gene synthesis

The oligonucleotides for the gene assembly were designed with DNAWorks v2.3 [100]. The amino acid sequence of human Delta-like protein 4 cytoplasmic region (DLL4_IC, corresponding to residues 553-685 of DLL4_HUMAN) was backtranslated using the *E. coli* Class II codon usage [101], and the generated DNA sequence (dll4_ic) was divided into 18 partially overlapping nucleotides with a maximum length of 40 bases, a calculated annealing temperature (Tm) of 60ºC, a T_m range of 2.9ºC and a minimal overlap of 13 bases. The oligonucleotide sequences were designed to have the lowest propensity to form hairpins within each oligonucleotide, and to contain no repeats that might lead to mispriming in the polymerase chain reaction (PCR). Synthetic oligonucleotides were purchased from Sigma-Genosys (0.05 μmol scale) and, after being dissolved in equimolar concentration (200 nM), assembled by PCR using *Pfu* polymerase (Promega) with the following forward and reverse primers (MWG-biotech, 0.05 μmol scale) containing the wanted restriction sites: 5'-TAA TAG TAG CAT ATG AAA CAC CAT CAC CAT CAC CAT CGC CAG CTG CGT CTG CGT-3' (the underlined sequence encodes the start methionine, followed by a lysine residue and a six-histidine tag) and 5'-TAG TAG GGA TCC TCA TTA AAC TTC AGT TGC GAT CAC GCA CTC ATT ACG TTC-3', respectively. PCR conditions were: 5 min at 95°C (hot start), 25 cycles of amplification (30 s denaturation at 95°C, 30 s annealing at 58°C, 90 sec elongation at 72°C), 10 min at

72 °C for the final elongation. The assembled and amplified synthetic gene, resulting in a sharp band of the correct size in the agarose gel, was digested by Nde I/BamH I, and ligated into a pET11a vector using standard procedures. DH5α *E. coli* cells were transformed with the dll4_ic-pET11a construct and selected on LB plates with 100 μg/mL ampicillin. The positive clones were sequenced by automatic DNA sequencing in both forward and reverse directions and the correct one (1/12) used for protein expression.

The nucleotide sequence of the truncated form of DLL4_IC lacking the first 23 amino acids (ΔN-DLL4_IC) was amplified by PCR from the dll4_ic-pET11a construct with the reverse primer used in the DLL4_IC gene synthesis and the following forward primer: 5'-TAG TAG TAG CAT ATG AAA GAT AAC CTG ATT CCG-3'. PCR conditions were the same as above except for the annealing temperature which was set at 58°C. The PCR product was digested by Nde I/BamH I, ligated into a pET11a vector, and the construct used to transform DH5α *E. coli* cells. The positive clones were selected as above, sequenced in both directions, and the correct one (1/2) used for protein expression.

Protein expression and purification

1) DLL4_IC

Purification in denaturing conditions. 1 L of LB containing 100 μg/mL ampicillin and 25 μg/mL chloramphenicol was inoculated with a clone of BL21(DE3)pLysS cells transformed with the dll4_ic-pET11a construct. Cells were grown at 37°C to an OD of ~0.8 and protein expression was induced with IPTG 1 mM for 3 h at room temperature. Cells were harvested, washed, and resuspended in the lysis buffer (20 mM phosphate buffer, 0.5 M NaCl, 50 mM CHAPS, 2% TWEEN, 5 mM TCEP, protein inhibitor cocktail tablet (Roche), 10 mM imidazole, 6 M GuHCl, pH 7.4) and sonicated on ice. After centrifugation and filtration through a 0.22 μm filter, the supernatant was loaded on a Ni^{2+} Sepharose His-Trap HP column (1 mL, Amersham Biosciences), the column washed with 20 mM phosphate buffer, 0.5 M NaCl, 10 mM imidazole, 5 mM TCEP, 6 M GuHCl, pH 7.4, and the protein eluted with a 0.01-0.5 M imidazole gradient. In order to remove fragments derived from partial proteolytic degradation, the eluted material was

purified by RP-HPLC with a Zorbax 300SB-CN column (9.4 x 250 mm, 5 µm, Agilent) using a 0-40% gradient of 0.1% TFA in H_2O and 0.1% TFA in CH_3CN and freeze-dried. The N-terminal His_6-tag was removed using a recombinant dipeptidyl aminopeptidase I (DAPase) containing a C-terminal His-tag (TAGzyme, Qiagen) for 2 h at 37ºC according to the manufacturer's protocol. An additional IMAC step on a His-Trap HP column (1 mL) using a 0-0.5 M imidazole gradient removed the peptidase, the partially digested protein and the cleaved His_2 dipeptides. The protein was subjected to a final RP-HPLC step and analyzed by LC-MS on a Gilson HPLC system coupled to an ESI-MS single quadrupole mass spectrometer (Applied Biosystems API-150EX), using a Zorbax 300SB-CN column (2.1 X 150 mm, 5 µm, Agilent) and a 0-50% gradient of 0.1% TFA in H_2O and 0.1% TFA in CH_3CN. Deconvolution of the multicharge ion spectrum was carried out using the BioMultiView software (Applied Biosystems). and confirmed the correct molecular size of the purified product (Mr calculated: 14894 Da; Mr observed: 14893 Da). The purified protein was freeze-dried and used for spectroscopic studies. The yield was ~8 mg protein per 1 L of culture.

Purification in native conditions. In order to test the possibility that the above harsh purification conditions (i.e., 6 M GuHCl in lysis and IMAC buffers, RP-HPLC acidic buffers [pH~2], freeze-drying process) could irreversibly denature the protein, DLL4_IC was also purified in native conditions. Cells were grown and protein expression was induced as described above. Cells were then harvested, washed and resuspended in the lysis buffer (20 mM phosphate buffer, 0.5 M NaCl, 50 mM CHAPS, 2% TWEEN, 5 mM TCEP, protein inhibitor cocktail tablet, 10 mM imidazole, pH 7.4). After the cells had been sonicated and spun as described, the supernatant was loaded onto a His-Trap HP column, which was washed with 20 mM phosphate buffer, 0.5 M NaCl, 10 mM imidazole, 5 mM TCEP, pH 7.4, and the protein eluted with a 0.01-0.5 M imidazole gradient. The eluted fractions were pooled and diluted with buffer A (20 mM phosphate buffer, 5 mM DTT, pH 7.4). The His_6-DLL4_IC protein was purifed by ion-exchange chromatography on an 8 x 75 mm SP column (SP-825, Shodex) using a 0-50% gradient from buffer A to buffer B (20 mM sodium phosphate, 1 M NaCl, 5 mM DTT, pH 7.4). The eluate was concentrated by ultrafiltration on a Centricon 3000 (Amicon), followed

by dilution with the final buffer (5 mM phosphate, 1 mM TCEP) for the subsequent spectroscopic analysis. Although the protein yield was quite low, the achieved protein purity was satisfactory (>95%) and the correct molecular size was confirmed by LC-MS.

^{15}N isotopic enrichment. 1 L of ^{15}N-M9 minimal medium (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1.3 g/L glycerol, 0.5 g/L $^{15}NH_4Cl$, 0.12 g/L $MgSO_4$, 0.01 g/L $CaCl_2$) containing nutrients supplemented as yeast nitrogen base w/o amino acids or NH_4SO_4 (1.7 g/L), pH 7, with 100 μg/mL ampicillin and 25 μg/mL chloramphenicol was inoculated with a clone of BL21(DE3)pLysS cells transformed with the dll4_ic-pET11a construct. Cells were grown overnight at room temperature to an OD of ~0.6 and protein expression was induced with IPTG (1 mM) for 4.5 h at 37ºC. Protein purification was carried out in denaturing conditions as described above, with a final yield of ~8 mg of pure product per 1 L of culture

Expression of ΔN-DLL4_IC and purification from inclusion bodies. 1 L of LB containing 100 μg/mL ampicillin and 25 μg/mL chloramphenicol was inoculated with a clone of BL21(DE3)pLysS cells transformed with the ΔN-dll4_IC construct. Cells were grown at 37°C to an OD of ~0.9 and protein expression induced with IPTG (1 mM) for 3 h at 37°C. Cells were harvested, washed and resuspended in the lysis buffer (50 mM Tris-HCl buffer, 5 mM EDTA, 0.5% Triton-X100, 0.1 mM PMSF, 1 mM DTT and protein inhibitor cocktail tablet (Roche)) and sonicated on ice. After sonication, $MgSO_4$ (10 mM) was added to chelate EDTA and the inclusion bodies collected by centrifugation at 6000 rpm for 15 min. The pellet was washed twice with lysis buffer and an additional wash was carried out without Triton-X100. The final inclusion body pellet was resuspended in 100 mM Tris-HCl, 50 mM Glycine, pH 8.0, dispersed by sonication and dissolved dropwise with the same buffer containing urea to a final concentration of 6 M urea. The urea was eliminated from the solution with a HiPrep 26/10 Desalting column and a final purification step was performed using RP-HPLC with a Zorbax 300SB-CN column (9.4 x 250 mm, 5 μm, Agilent) using a 0-60% gradient of 0.1% TFA in H_2O and 0.1% TFA in CH_3CN. The freeze-dried product was analyzed by LC-MS which confirmed the correct

molecular size (Mr calculated: 1267.9 Da; Mr observed: 1265.0 Da).The yield was ~4 mg protein (~95% purity) per 1 L of culture.

2) PDZ1

The DNA encoding the first PDZ domain of rat Dlg-1 (PDZ1, residues 221–311), was amplified by PCR from the c-DNA of the full length Dlg-1 mutant used in the *in vitro* binding assay, with the following forward and reverse primers (MWG-biotech, 0.05 μmol scale): 5'-TAG TAG CAT ATG GAA TAT GAA GAA ATC ACA-3' and 5'-TAG TAG GGA TCC TCA TTA ATG GTG ATG GTG ATG GTG TTT CCT TCT TTT TAC ATA-3' (the underlined sequence encodes a six-histidine tag), respectively. PCR conditions were: 5 min at 95°C (hot start), 25 cycles of amplification (30 s denaturation at 95°C, 30 s annealing at 36°C, 90 sec elongation at 72°C), 10 min at 72 °C for the final elongation. The PCR product was cloned into the Nde I/BamHI sites of a pET11a vector.
1 L of LB containing 100 μg/mL ampicillin was inoculated with a clone of BL21(DE3) cells transformed with the PDZ1-pET11a construct. Cells were grown at 37°C to an OD of ~0.8 and protein expression was induced with IPTG 1 mM for 3 h at room temperature. Cells were harvested, washed, and resuspended in the lysis buffer (20 mM phosphate buffer, 0.5 M NaCl, 50 mM CHAPS, 2% TWEEN, 5 mM TCEP, protein inhibitor cocktail tablet (Roche), 10 mM imidazole, pH 7.4) and sonicated on ice. After centrifugation and filtration through a 0.22 μm filter, the supernatant was loaded on a Ni^{2+} Sepharose His-Trap HP column (1 mL, Amersham Biosciences), the column washed with 20 mM phosphate buffer, 0.5 M NaCl, 10 mM imidazole, 5 mM TCEP, pH 7.4, and the protein eluted with a 0.01-0.5 M imidazole gradient. Imidazole was removed with a HiPrep 26/10 Desalting column (Amersham) with a desalting buffer (5 mM Tris, 1 mM TCEP, pH 7.4 or 20 mM phosphate buffer, 1mM TCEP, pH 7.4) and the sample concentrated with a 15 ml Centriprep column (Amicon). The ^{15}N isotopic enrichment was carried out as described for DLL4_IC, by substituting glycerol with glucose (2g/L) in the minimal medium. The desalted and concentrated PDZ1 protein was used for the far-UV CD and NMR analyses.

Peptide synthesis

All peptides (P1, res. 582-618; P2, res. 619-661; P3, res. 662-685) were prepared by standard solid-phase Fmoc methods using a home-built automatic synthesizer based on a Gilson Aspec XL SPE. After cleavage/deprotection of the peptide-resin (preloaded NovaSyn TGT, Novabiochem) in TFA/1,2-ethanedithiol/triisopropylsilane/H_2O 90/5/2.5/2.5 v/v/v for 2 h, deprotected, reduced peptides were purified by semi-preparative RP-HPLC on a Zorbax 300SB-C18 column (9.4 x 250 mm, 5 µm, Agilent) and freeze-dried. P1 was purified using a 0-60% gradient of 0.1% TFA in H_2O and 0.1% TFA in CH_3CN. P2 and P3 were purified using a 0-40% gradient of triethylammonium acetate (TEAA) (10 mM, pH 7 in H_2O) and TEAA in 80% CH_3CN, followed by further purification and desalting using a 0-60% gradient of 0.1% TFA in H_2O and 0.1% TFA in CH_3CN. The identity of the peptides was checked by LC-MS and the yield and purity estimated from RP-HPLC. Final yields were in the range 40–60% and purity > 95%.

Size exclusion chromatography

The freeze-dried protein powder was dissolved in the elution buffer (50 mM Tris-HCl, 100 mM KCl, pH 7.4), loaded onto a Sephacryl S-200 column (Pharmacia) and eluted in the same elution buffer. The apparent molecular mass of DLL4_IC was deduced from a calibration carried out with the following molecular standards: lactate dehydrogenase (147 kDa), bovine serum albumin (67 kDa), carbonic anhydrase (29 kDa) and horse myoglobin (17 kDa). Stokes radii of native (R_SN) and fully unfolded (R_SU) proteins of known molecular weight (MW) were determined according to the equations described by Uversky [62]:

$\log(R_SN) = -(0.254 \pm 0.002) + (0.369 \pm 0.001) \log(MW)$, and

$\log(R_SU) = -(0.543 \pm 0.004) + (0.502 \pm 0.001) \log(MW)$.

Circular dichroism (CD)

The freeze-dried DLL4_IC protein powder was dissolved either in 5 mM MES buffer, 1 mM TCEP, pH 6.3, or in 5 mM Tris buffer, 1 mM TCEP, pH 7.5. Protein concentration was determined by UV absorbance at 280 nm using the calculated ε value of 2560 $M^{-1}cm^{-1}$. CD spectra of solutions of DLL4_IC, ΔN-DLL4_IC, PDZ1 or the synthetic peptides, were recorded on a Jasco-810 spectropolarimeter in the 190-250 nm range, using quartz cuvettes (path length 0.1 cm). Spectra were averaged from 5 scans of 0.1 nm steps at 20-50 nm/min. The secondary structure analysis was performed using the Dichroweb [102] tool CDSSTR. The helical content was also determined from the mean residue ellipticity (MRE, deg·cm^2/dmol) at 222 nm ($[\theta]_{222}$) according to the equation: $[\alpha] = 100\cdot[\theta]_{222}/\theta_f$ and $\theta_f = -40000\cdot(1-2.57/n)$ where $[\alpha]$ is the amount of helix, *n* is the number of residues, and θ_f is the maximum MRE of an α-helix of *n* residues [103] .

NMR spectroscopy

The sample for NMR spectroscopy was prepared by dissolving the freeze-dried material in H_2O/D_2O (90/10, v/v) containing 4 mM TCEP, 2 mM EDTA-d_{16}, 15 μM DSS and adjusting the pH to 5.6 with small aliquots of 0.1 N NaOH, for a final protein concentration of ~0.5 mM. The sample containing SDS was prepared dissolving solid SDS sodium salt in the NMR sample, for a final SDS concentration of 50 mM. After flushing the NMR tube with argon, spectra were recorded at 303 K on a Bruker spectrometer operating at a ^1H frequency of 600.13 MHz and equipped with a $^1H/^{13}C/^{15}N$ triple resonance Z-axis gradient probe. Transmitter frequencies in the ^1H and ^{15}N dimensions were set on the water line and at 118.0 ppm, respectively. HSQC and HSQC-TOCSY experiments were carried out in phase-sensitive mode using echo/antiecho-TPPI gradient selection and ^{15}N decoupling during acquisition. HSQC spectra were acquired with 1K complex points, 256 t_1 experiments, 32 scans per increment, over a spectral width of 13 and 28 ppm in the ^1H and ^{15}N dimensions, respectively. HSQC-TOCSY spectra were acquired with the same parameters, but with 128 scans per t_1 increment and a 40 ms DIPSI mixing time. Data were transformed using X-WinNMR (Bruker) and

analyzed using CARA (http://www.nmr.ch). ^{1}H chemical shifts were referenced to internal DSS.

In vitro binding assay

The rat Dlg-1-deleted mutant derivatives are a kind gift from Lawrence Banks [104]. The constructs, confirmed by partial sequence analysis, were used for the *in vitro* expression of Dlg proteins. *In vitro* binding assays were performed in the following fashion: *in vitro*-translated, ^{35}S radiolabeled Dlg proteins were incubated at 4 °C for 2 hrs in the presence of nickel-sepharose beads to which a recombinant, purified His_6-DLL4_IC protein was previously conjugated, in the binding buffer (20 mM phosphate buffer, 0.5 M NaCl, 50 mM imidazole, 5 mM TCEP, protein inhibitor cocktail tablet (Roche), pH 7.4). As a negative control, the Dlg proteins were separately incubated with unconjugated beads. After extensive washing with the washing buffer (binding buffer added with 1% Tween), the beads were collected and the Dlg proteins were eluted from the beads with SDS sample buffer, run on a SDS-PAGE gel (12%) and assayed by autoradiography.

Dataset preparation and analysis

A set of human membrane proteins was generated by a search of the Swiss-Prot database through the Sequence Retrieval System using “receptor” and “transmembrane” as keywords and “single-pass” in the comment field. Entries having type II topology were manually discarded from the dataset. The total number of sequences thereby collected was 369. This dataset was then divided into two subsets containing intracellular and extracellular domains named “intracellular subset” and “extracellular subset”, respectively. The boundaries of the domains were selected according to the position of the transmembrane helix in the sequences as provided in Swiss-Prot. The DisProt [78] dataset (release 3.6) was downloaded from the Database of Protein Disorder (www.disprot.com) and contained 469 entries, while the reduced SCOP dataset was created from the SCOP [79] database 1.69 (http://astral.berkeley.edu) by discarding all entries with >40% identity, and contained 1357 sequences. Disorder predictions were

carried out using DisEMBL [80], IUPred [60], charge/hydropathy plots [81], and amino acid compositional analysis [105]. DisEMBL (v. 1.5; http://dis.embl.de) was run using the three definitions of protein disorder, based on assignments of secondary structure (loops/coils), high values of Cα B-factors (hot loops), and missing coordinates in X-Ray structures (Remark465). IUPred (http://iupred.enzim.hu/) was run calculating pairwise energies within a window of 100 or 25 residues ("long" and "short" disorder definitions, respectively). Charge/hydropathy for each sequence was obtained from the absolute value of the mean net charge versus the mean residue hydropathy calculated using the normalized Kyte-Doolittle scale. Amino acid compositional analysis was carried out using Composition Profiler [106] (http://www.cprofiler.org) using the PDB Select 25 [107] or the DisProt [78] datasets as reference for ordered and disordered proteins, respectively. Enrichment or depletion in each amino acid type was expressed as $(C_{s1}-C_{s2})/C_{s2}$, i.e., the normalized excess of a given residue's "concentration" in a dataset (C_{s1}) relative to the corresponding value in the other dataset (C_{s2}). Amino acid types were ranked according to increasing flexibility [108].

Disorder predictors used

PONDR [56] (Predictor Of Naturally Disordered Regions) predictors are feedforward neural networks that use sequence information from windows of generally 21 amino acids. Attributes, such as the fractional composition of particular amino acids or hydropathy, are calculated over this window, and these values are used as inputs for the predictor. The neural network, which has been trained on a specific set of ordered and disordered sequences, then outputs a value for the central amino acid in the window.

DisEMBL [77] is a neural network-based predictor of disorder that uses three definitions of disorder (loops/coils, hot loops or Remark465) as described in the "Dataset preparation" paragraph. For each of the three definitions of disorder, a data set was constructed and the neural networks trained on each of the three data sets.

IUPred [57] is a predictor that discriminates between ordered and disordered regions based on the potential of polypeptides to form stabilizing interresidue interactions, like those observed in globular proteins. As the sum of the interaction energies can be approximated

by a quadratic expression in the amino acid composition, it is possible to estimate the pairwise interaction energies using a set of globular proteins of known structure. The calculation involves a 20 X 20 energy predictor matrix, parameterized by a statistical method to approach the expected pairwise energy of the globular set. These interactions energies can then be used to estimate the energy content of a structurally uncharacterized or intrinsically disordered protein region based only on its amino acid composition. Comparing globular proteins and disordered ones, a clear separation in their energy content is observed. The approach is turned into a position-specific predictor of protein disorder that considers only the local (± 2-100 residues) sequential environment of residues. The score is averaged over a 21 residue window. IUPred does not rely on networks trained with datasets of disordered proteins, which can be limited in size and heterogeneous in terms of experimental conditions; the propensity to disorder estimated by IUPred reflects an intrinsic property of certain proteins: the inability to form a number of favorable interresidue interactions sufficient to overcome the entropy loss during folding.

References

1. Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. *Science*. 1999;284(5415):770-6.

2. Kadesch T. Notch signaling: the demise of elegant simplicity. *Curr Opin Genet Dev*. 2004;14(5):506-12.

3. Nam Y, Aster JC, Blacklow SC. Notch signaling as a therapeutic target. *Curr Opin Chem Biol*. 2002;6(4):501-9.

4. Letunic I, Copley RR, Pils B, et al. SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res*. 2006;34(Database issue):D257-60.

5. Rebay I, Fleming RJ, Fehon RG, et al. Specific EGF repeats of Notch mediate interactions with Delta and Serrate: implications for Notch as a multifunctional receptor. *Cell*. 1991;67(4):687-99.

6. Shimizu K, Chiba S, Kumano K, et al. Mouse jagged1 physically interacts with notch2 and other notch receptors. Assessment by quantitative methods. *J Biol Chem*. 1999;274(46):32961-9.

7. Weinmaster G. Notch signal transduction: a real rip and more. *Curr Opin Genet Dev*. 2000;10(4):363-9.

8. Haines N, Irvine KD. Glycosylation regulates Notch signalling. *Nat Rev Mol Cell Biol*. 2003;4(10):786-97.

9. Lai EC. Keeping a good pathway down: transcriptional repression of Notch pathway target genes by CSL proteins. *EMBO Rep*. 2002;3(9):840-5.

10. Allman D, Punt JA, Izon DJ, Aster JC, Pear WS. An invitation to T and more: notch signaling in lymphopoiesis. *Cell*. 2002;109 Suppl:S1-11.

11. Radtke F, Raj K. The role of Notch in tumorigenesis: oncogene or tumour suppressor? *Nat Rev Cancer*. 2003;3(10):756-67.

12. Weng AP, Ferrando AA, Lee W, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science*. 2004;306(5694):269-71.

13. Xue Y, Gao X, Lindsell CE, et al. Embryonic lethality and vascular defects in mice lacking the Notch ligand Jagged1. *Hum Mol Genet*. 1999;8(5):723-30.

14. Hrabě de Angelis M, McIntyre J, Gossler A. Maintenance of somite borders in mice requires the Delta homologue DII1. *Nature*. 1997;386(6626):717-21.

15. Gale NW, Dominguez MG, Noguera I, et al. Haploinsufficiency of delta-like 4 ligand results in embryonic lethality due to major defects in arterial and vascular development. *Proc Natl Acad Sci U S A*. 2004;101(45):15949-54.

16. Jiang R, Lan Y, Chapman HD, et al. Defects in limb, craniofacial, and thymic development in Jagged2 mutant mice. *Genes Dev*. 1998;12(7):1046-57.

17. Thurston G, Noguera-Troise I, Yancopoulos GD. The Delta paradox: DLL4 blockade leads to more tumour vessels but less tumour growth. *Nat Rev Cancer*. 2007;7(5):327-31.

18. Siekmann AF, Lawson ND. Notch signalling limits angiogenic cell behaviour in developing zebrafish arteries. *Nature*. 2007;445(7129):781-4.

19. Suchting S, Freitas C, le Noble F, et al. The Notch ligand Delta-like 4 negatively regulates endothelial tip cell formation and vessel branching. *Proc Natl Acad Sci U S A*. 2007;104(9):3225-30.

20. Scehnet JS, Jiang W, Kumar SR, et al. Inhibition of Dll4-mediated signaling induces proliferation of immature vessels and results in poor tissue perfusion. *Blood*. 2007;109(11):4753-60.

21. Lobov IB, Renard RA, Papadopoulos N, et al. Delta-like ligand 4 (Dll4) is induced by VEGF as a negative regulator of angiogenic sprouting. *Proc Natl Acad Sci U S A*. 2007;104(9):3219-24.

22. Leslie JD, Ariza-McNaughton L, Bermange AL, et al. Endothelial signalling by the Notch ligand Delta-like 4 restricts angiogenesis. *Development*. 2007;134(5):839-44.

23. Hellström M, Phng L, Hofmann JJ, et al. Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis. *Nature*. 2007;445(7129):776-80.

24. Ridgway J, Zhang G, Wu Y, et al. Inhibition of Dll4 signalling inhibits tumour growth by deregulating angiogenesis. *Nature*. 2006;444(7122):1083-7.

25. Noguera-Troise I, Daly C, Papadopoulos NJ, et al. Blockade of Dll4 inhibits tumour growth by promoting non-productive angiogenesis. *Nature*. 2006;444(7122):1032-7.

26. Yan M, Plowman GD. Delta-like 4/Notch Signaling and Its Therapeutic Implications. *Clin Cancer Res*. 2007;13(24):7243-6.

27. Sainson RCA, Harris AL. Anti-Dll4 therapy: can we block tumour growth by increasing angiogenesis? *Trends Mol Med*. 2007;13(9):389-95.

28. Pintar A, De Biasio A, Popovic M, Ivanova N, Pongor S. The intracellular region of Notch ligands: does the tail make the difference? *Biol Direct*. 2007;2:19.

29. Hock B, Böhme B, Karn T, et al. PDZ-domain-mediated interaction of the Eph-related receptor tyrosine kinase EphB3 and the ras-binding protein AF6 depends on the kinase activity of the receptor. *Proc Natl Acad Sci U S A*. 1998;95(17):9779-84.

30. Six EM, Ndiaye D, Sauer G, et al. The notch ligand Delta1 recruits Dlg1 at cell-cell contacts and regulates cell migration. *J Biol Chem*. 2004;279(53):55818-26.

31. Wright GJ, Leslie JD, Ariza-McNaughton L, Lewis J. Delta proteins and MAGI proteins: an interaction of Notch ligands with intracellular scaffolding molecules and its significance for zebrafish development. *Development*. 2004;131(22):5659-69.

32. Pfister S, Przemeck GKH, Gerber J, et al. Interaction of the MAGUK family member Acvrinp1 and the cytoplasmic domain of the Notch ligand Delta1. *J Mol Biol*. 2003;333(2):229-35.

33. Beuming T, Skrabanek L, Niv MY, Mukherjee P, Weinstein H. PDZBase: a protein-protein interaction database for PDZ-domains. *Bioinformatics*. 2005;21(6):827-8.

34. Basdevant N, Weinstein H, Ceruso M. Thermodynamic basis for promiscuity and selectivity in protein-protein interactions: PDZ domains, a case study. *J Am Chem Soc*. 2006;128(39):12766-77.

35. Wiedemann U, Boisguerin P, Leben R, et al. Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. *J Mol Biol*. 2004;343(3):703-18.

36. Kanwar R, Fortini ME. Notch signaling: a different sort makes the cut. *Curr Biol*. 2004;14(24):R1043-5.

37. Le Borgne R, Schweisguth F. Notch signaling: endocytosis makes delta signal better. *Curr Biol*. 2003;13(7):R273-5.

38. Le Borgne R, Bardin A, Schweisguth F. The roles of receptor and ligand endocytosis in regulating Notch signaling. *Development*. 2005;132(8):1751-62.

39. Nichols JT, Miyamoto A, Olsen SL, et al. DSL ligand endocytosis physically dissociates Notch1 heterodimers before activating proteolysis can occur. *J Cell Biol*. 2007;176(4):445-58.

40. Itoh M, Kim C, Palardy G, et al. Mind bomb is a ubiquitin ligase that is essential for efficient activation of Notch signaling by Delta. *Dev Cell*. 2003;4(1):67-82.

41. Wang W, Struhl G. Drosophila Epsin mediates a select endocytic pathway that DSL ligands must enter to activate Notch. *Development*. 2004;131(21):5367-80.

42. Takeuchi T, Adachi Y, Ohtsuki Y. Skeletrophin, a novel ubiquitin ligase to the intracellular region of Jagged-2, is aberrantly expressed in multiple myeloma. *Am J Pathol*. 2005;166(6):1817-26.

43. Song R, Koo B, Yoon K, et al. Neuralized-2 regulates a Notch ligand in cooperation with Mind bomb-1. *J Biol Chem*. 2006;281(47):36391-400.

44. Ascano JM, Beverly LJ, Capobianco AJ. The C-terminal PDZ-ligand of JAGGED1 is essential for cellular transformation. *J Biol Chem*. 2003;278(10):8771-9.

45. Ikeuchi T, Sisodia SS. The Notch ligands, Delta1 and Jagged2, are substrates for presenilin-dependent "gamma-secretase" cleavage. *J Biol Chem*. 2003;278(10):7751-4.

46. LaVoie MJ, Selkoe DJ. The Notch ligands, Jagged and Delta, are sequentially processed by alpha-secretase and presenilin/gamma-secretase and release signaling fragments. *J Biol Chem*. 2003;278(36):34427-37.

47. Six E, Ndiaye D, Laabi Y, et al. The Notch ligand Delta1 is sequentially cleaved by an ADAM protease and gamma-secretase. *Proc Natl Acad Sci U S A*. 2003;100(13):7638-43.

48. Hiratochi M, Nagase H, Kuramochi Y, et al. The Delta intracellular domain mediates TGF-beta/Activin signaling through binding to Smads and has an important bi-directional function in the Notch-Delta signaling pathway. *Nucleic Acids Res*. 2007;35(3):912-22.

49. Vardar D, North CL, Sanchez-Irizarry C, Aster JC, Blacklow SC. Nuclear magnetic resonance structure of a prototype Lin12-Notch repeat module from human Notch1. *Biochemistry*. 2003;42(23):7061-7.

50. Hambleton S, Valeyev NV, Muranyi A, et al. Structural and functional properties of the human notch-1 ligand binding region. *Structure*. 2004;12(12):2173-83.

51. Ehebauer MT, Chirgadze DY, Hayward P, Martinez Arias A, Blundell TL. High-resolution crystal structure of the human Notch 1 ankyrin domain. *Biochem J*. 2005;392(Pt 1):13-20.

52. Lubman OY, Kopan R, Waksman G, Korolev S. The crystal structure of a partial mouse Notch-1 ankyrin domain: repeats 4 through 7 preserve an ankyrin fold. *Protein Sci*. 2005;14(5):1274-81.

53. Kovall RA, Hendrickson WA. Crystal structure of the nuclear effector of Notch signaling, CSL, bound to DNA. *EMBO J*. 2004;23(17):3441-51.

54. Popovic M, Coglievina M, Guarnaccia C, et al. Gene synthesis, expression, purification, and characterization of human Jagged-1 intracellular region. *Protein Expr Purif*. 2006;47(2):398-404.

55. Popovic M, De Biasio A, Pintar A, Pongor S. The intracellular region of the Notch ligand Jagged-1 gains partial structure upon binding to synthetic membranes. *FEBS J*. 2007;274(20):5325-36.

56. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*. 1999;292(2):195-202.

57. Cuff JA, Barton GJ. Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins*. 1999;34(4):508-19.

58. Pollastri G, Przybylski D, Rost B, Baldi P. Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins*. 2002;47(2):228-35.

59. Obradovic Z, Peng K, Vucetic S, et al. Predicting intrinsic disorder from amino acid sequence. *Proteins*. 2003;53 Suppl 6:566-72.

60. Dosztányi Z, Csizmók V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol*. 2005;347(4):827-39.

61. Wishart DS, Sykes BD, Richards FM. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol*. 1991;222(2):311-33.

62. Uversky VN. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry*. 1993;32(48):13288-98.

63. Roccatano D, Colombo G, Fioroni M, Mark AE. Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: a molecular dynamics study. *Proc Natl Acad Sci U S A*. 2002;99(19):12179-84.

64. Chandra S, Chen X, Rizo J, Jahn R, Südhof TC. A broken alpha -helix in folded alpha -Synuclein. *J Biol Chem*. 2003;278(17):15313-8.

65. Wang Y, Jardetzky O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci*. 2002;11(4):852-61.

66. Tsunoda S, Sierralta J, Zuker CS. Specificity in signaling pathways: assembly into multimolecular signaling complexes. *Curr Opin Genet Dev*. 1998;8(4):419-22.

67. Kennedy MB. Origin of PDZ (DHR, GLGF) domains. *Trends Biochem Sci*. 1995;20(9):350.

68. Harris BZ, Lim WA. Mechanism and role of PDZ domains in signaling complex assembly. *J Cell Sci*. 2001;114(Pt 18):3219-31.

69. Kornau HC, Schenker LT, Kennedy MB, Seeburg PH. Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. *Science*. 1995;269(5231):1737-40.

70. Harrison SC. Peptide-surface association: the case of PDZ and PTB domains. *Cell*. 1996;86(3):341-3.

71. Doyle DA, Lee A, Lewis J, et al. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell*. 1996;85(7):1067-76.

72. Jeleń F, Oleksy A, Smietana K, Otlewski J. PDZ domains - common players in the cell signaling. *Acta Biochim Pol*. 2003;50(4):985-1017.

73. Songyang Z, Fanning AS, Fu C, et al. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science*. 1997;275(5296):73-7.

74. Feng W, Long J, Fan J, Suetake T, Zhang M. The tetrameric L27 domain complex as an organization platform for supramolecular assemblies. *Nat Struct Mol Biol*. 2004;11(5):475-80.

75. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. *Proteins*. 1995;23(4):566-79.

76. Wang L, Piserchio A, Mierke DF. Structural characterization of the intermolecular interactions of synapse-associated protein-97 with the NR2B subunit of N-methyl-D-aspartate receptors. *J Biol Chem*. 2005;280(29):26992-6.

77. el-Agnaf OMA, Irvine GB. Aggregation and neurotoxicity of alpha-synuclein and related peptides. *Biochem Soc Trans*. 2002;30(4):559-65.

78. Sickmeier M, Hamilton JA, LeGall T, et al. DisProt: the Database of Disordered Proteins. *Nucleic Acids Res*. 2007;35(Database issue):D786-93.

79. Andreeva A, Howorth D, Chandonia J, et al. Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res*. 2008;36(Database issue):D419-25.

80. Linding R, Jensen LJ, Diella F, et al. Protein disorder prediction: implications for structural proteomics. *Structure*. 2003;11(11):1453-9.

81. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*. 2000;41(3):415-27.

82. Uversky VN. What does it mean to be natively unfolded? *Eur J Biochem*. 2002;269(1):2-12.

83. Dunker AK, Lawson JD, Brown CJ, et al. Intrinsically disordered protein. *J Mol Graph Model*. 2001;19(1):26-59.

84. Tompa P, Szász C, Buday L. Structural disorder throws new light on moonlighting. *Trends Biochem Sci*. 2005;30(9):484-9.

85. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*. 2005;6(3):197-208.

86. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*. 2004;337(3):635-45.

87. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol*. 2002;323(3):573-84.

88. Mohan A, Oldfield CJ, Radivojac P, et al. Analysis of molecular recognition features (MoRFs). *J Mol Biol*. 2006;362(5):1043-59.

89. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters*. 2005;579(15):3346-3354.

90. Olivotto M, Arcangeli A, Carlà M, Wanke E. Electric fields at the plasma membrane level: a neglected element in the mechanisms of cell signalling. *Bioessays*. 1996;18(6):495-504.

91. Slavík J. Intracellular pH topography: determination by a fluorescent probe. *FEBS Lett*. 1983;156(2):227-30.

92. Brown CJ, Takayama S, Campen AM, et al. Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol*. 2002;55(1):104-10.

93. Tompa P. Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays*. 2003;25(9):847-55.

94. Sipos L, von Heijne G. Predicting the topology of eukaryotic membrane proteins. *Eur J Biochem*. 1993;213(3):1333-40.

95. Sigalov AB, Aivazian DA, Uversky VN, Stern LJ. Lipid-binding activity of intrinsically unstructured cytoplasmic domains of multichain immune recognition receptor signaling subunits. *Biochemistry*. 2006;45(51):15731-9.

96. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*. 1999;294(5):1351-62.

97. Puntervoll P, Linding R, Gemünd C, et al. ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res*. 2003;31(13):3625-30.

98. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput*. 2002:310-22.

99. Catic A, Collins C, Church GM, Ploegh HL. Preferred in vivo ubiquitination sites. *Bioinformatics*. 2004;20(18):3302-7.

100. Hoover DM, Lubkowski J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res*. 2002;30(10):e43.

101. Médigue C, Rouxel T, Vigier P, Hénaut A, Danchin A. Evidence for horizontal gene transfer in Escherichia coli speciation. *J Mol Biol*. 1991;222(4):851-6.

102. Lobley A, Whitmore L, Wallace BA. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics*. 2002;18(1):211-2.

103. Chen YH, Yang JT, Chau KH. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry*. 1974;13(16):3350-9.

104. Gardiol D, Galizzi S, Banks L. Mutational analysis of the discs large tumour suppressor identifies domains responsible for human papillomavirus type 18 E6-mediated degradation. *J Gen Virol*. 2002;83(Pt 2):283-9.

105. Romero P, Obradovic Z, Li X, et al. Sequence complexity of disordered protein. *Proteins*. 2001;42(1):38-48.

106. Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics*. 2007;8:211.

107. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-42.

108. Vihinen M, Torkkila E, Riikonen P. Accuracy of protein flexibility predictions. *Proteins*. 1994;19(2):141-9.