

Prevalence of Intrinsic Disorder in the Intracellular Region of Human Single-Pass Type I Proteins: The Case of the Notch Ligand Delta-4

Alfredo De Biasio,[§] Corrado Guarnaccia,[§] Matija Popovic,[§] Vladimir N. Uversky,^{†,‡}
 Alessandro Pintar,^{*,§} and Sándor Pongor^{*,§}

Protein Structure and Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), AREA Science Park, Padriciano 99, I-34012 Trieste, Italy, Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia, and Institute for Intrinsically Disordered Protein Research, Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 410 W. 10th Street, HS 5000, Indianapolis, Indiana 46202

Received January 25, 2008

Intrinsic disorder (ID) is a widespread phenomenon found especially in signaling and regulation-related eukaryotic proteins. The functional importance of flexible disordered regions often resides in their ability to allow proteins to bind different partners. The incidence and location of intrinsic disorder in 369 human single-pass transmembrane receptors with the type I topology was assessed based on both disorder predictions and amino acid physico-chemical properties. We provide evidence that ID concentrates in the receptors' cytoplasmic region. As a benchmark for this analysis, we present a structural study on the previously uncharacterized intracellular region of human Delta-4 (DLL4_IC), a single-pass transmembrane protein and a ligand of Notch receptors. DLL4_IC is required for receptor/ligand endocytosis; it undergoes regulated intramembrane proteolysis, and mediates protein–protein interactions through its C-terminal PDZ binding motif. Using a recombinant purified protein, we demonstrate using various biophysical methods that DLL4_IC is mainly disordered in solution but can form interconvertible local secondary structures in response to variations in the physico-chemical milieu. Most of these conformational changes occur in the highly conserved C-terminal segment that includes the PDZ-binding motif. On the basis of our results, we propose that global disorder, in concert with local preorganization, may play a role in Notch signaling mediated by Delta-4.

Keywords: Intrinsic disorder • single-pass transmembrane receptors • Notch signaling • PDZ-binding motif • SDS micelles • TFE

Introduction

Intrinsic disorder (ID) 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.^{1,2} Extensive disorder predictions in genomes of increasing complexity revealed that ID is the highest in eukaryotes³ and that long ID regions are mostly found in proteins involved in regulation and signaling.⁴ These two observations were interpreted as the consequence of the increased complexity of the regulatory networks in multicellular organisms, where the plasticity of ID regions in highly connected proteins is exploited to bind different signaling partners through disorder-to-order transitions.⁵

Systematic analyses carried out on human transcription factors revealed that they have, in general, a high fraction of residues predicted to be disordered.^{6,7} A recent report computed the distribution of ID in plasma membrane proteins of different transmembrane classes.⁸ Intriguingly, the authors showed that, in general, ID is more prominent in the cytoplasmic segments as compared to the extracellular ones and proposed that the phenomenon reflects the different cellular niches of proteins (inside versus outside). However, the function of this skewed ID distribution remains unclear. In the report, ID regions in single-spanning membrane proteins appear to be evenly distributed on both sides of the membrane, in contrast to other transmembrane classes.

Here, we focus on the ID incidence and location (intra- or extracellular) in the subclass of human single-pass transmembrane proteins with the type I topology (i.e., proteins with a single transmembrane helix and the N-terminus located in the extracellular space) and annotated by the Swiss-Prot database as “receptor”. Our disorder predictions, mean net charge versus hydrophathy plots, and amino acid compositional analysis indicate that ID has a relevant incidence in this class of

* To whom correspondence should be addressed. E-mail: (A.P.) pintar@icgeb.org; (S.P.) pongor@icgeb.org.

[§] International Centre for Genetic Engineering and Biotechnology (ICGEB).

[†] Russian Academy of Sciences.

[‡] Indiana University School of Medicine.



Figure 1. Domain architecture of Delta-4. MNLL, N-terminal domain; DSL, Delta/Serrate/Lag-2 domain; EGF, Epidermal Growth Factor repeat; the transmembrane segment is shown as a blue bar.

proteins, and that it is mostly concentrated in the intracellular region. As a benchmark for this analysis, we present an experimental study on a relevant target, Delta-4, a single-pass transmembrane protein and a ligand of Notch receptors. The Notch pathway is a highly conserved, intercellular and bidirectional mechanism that regulates cell differentiation.^{9–11} Five different ligands of Notch receptors have been identified in mammals, three *Drosophila* Delta orthologs (Delta-like-1, -3, and -4) and two *Drosophila* Serrate orthologs (Jagged-1 and -2). In recent studies, it was shown that inhibition of the Notch pathway mediated by Delta-4 can promote tumor regression by deregulating angiogenesis, making Delta-4 a potential pharmacological target for the treatment of solid tumors.^{12,13} Delta-4 (Figure 1) is composed of an extracellular portion, consisting mainly of globular domains (i.e., a Delta/Serrate/Lag-2 (DSL) domain and eight epidermal growth factor-like repeats¹⁴), which mediates the binding to the receptor, a transmembrane segment, and a cytoplasmic tail of 133 amino acids which is likely to be involved in “receptor shedding” through ubiquitination and endocytosis.¹⁵ Also, the cytoplasmic tail of Notch ligands couples the Notch signaling network to PDZ-bearing proteins localized at the membrane/cytoplasm interface and is involved in the organization of cell–cell junctions. Delta-4 interacts through its conserved C-terminal motif (ATEV) with the PDZ domains of Dlg-1, the human homologue of *Drosophila* Discs Large protein.¹⁶ Finally, there is compelling evidence that, as a consequence of the proteolytic cleavage of the intracellular region of the ligand and its release from the cell membrane,^{17–19} Notch-related signal transduction is active not only in the signal-receiving cell, but also in the signal-sending one.^{17,20,21} The intracellular region of Delta-4 does not display any sequence homology with proteins of known fold. Therefore, our main concern here was to assess whether it might encode a new globular fold or, inversely, if it is partly or entirely disordered.

Our studies show that a recombinant protein representing the cytoplasmic tail of Delta-4 is intrinsically disordered in solution and that its C-terminal segment is highly plastic, as it can switch between three conformational states (coil, strand, helix) in response to changes in the chemical milieu (pH, dielectric permittivity, hydrophobic–hydrophilic interface). We discuss these results in light of our findings on the ID distribution in single-pass transmembrane proteins and propose that global disorder, in concert with local preorganization, may play a role in Notch signaling mediated by Delta-4.

Materials and Methods

Data Set 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 “recep-

tor” and “transmembrane” as keywords and “single-pass” in the comment field. Entries having type II topology were manually discarded from the data set. The total number of sequences thereby collected was 369. This data set 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 DisProt²² data set (release 3.6) was downloaded from the Database of Protein Disorder (www.disprot.com) and contained 469 entries, while the reduced SCOP data set was created from the SCOP²³ 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,²⁴ IUPred,²⁵ charge/hydrophathy plots,²⁶ and amino acid compositional analysis.²⁷ 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 α 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/hydrophathy for each sequence was obtained from the absolute value of the mean net charge versus the mean residue hydrophathy calculated using the normalized Kyte-Doolittle scale. Amino acid compositional analysis was carried out using Composition Profiler²⁸ (<http://www.cprofiler.org>) using the PDB Select 25²⁹ or the DisProt²² data sets as reference for ordered and disordered proteins, respectively. Enrichment or depletion in each amino acid type was expressed as $(C_{s1} - C_{s2})/C_{s2}$, that is, the normalized excess of a given residue’s “concentration” in a data set (C_{s1}) relative to the corresponding value in the other data set (C_{s2}). Amino acid types were ranked according to increasing flexibility.³⁰

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.³¹ Secondary structure predictions (PSIPRED,³² JNet,³³ SSpro³⁴) 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 v2.3.³⁵ 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 *Escherichia coli* Class II codon usage,³⁶ 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 (T_m) of 60 °C, a T_m range of 2.9 °C 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 μ mol 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 μ mol 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 s elongation at 72 °C), and 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. DH5 α *E. coli* cells were transformed with the dll4_ic-pET11a construct and selected on LB plates with 100 μ g/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 (Δ N-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 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 DH5 α *E. coli* cells. The positive clones were selected as above and sequenced in both directions, and the correct one (1/2) was used for protein expression.

Protein Expression and Purification. 1. Purification in Denaturing Conditions. One liter of LB containing 100 μ g/mL ampicillin and 25 μ g/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 \sim 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, and 6 M GuHCl, pH 7.4) and sonicated on ice. After centrifugation and filtration through a 0.22 μ m filter, the supernatant was loaded on a Ni²⁺-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, and 6 M GuHCl, pH 7.4, and the protein eluted with a 0.01–0.5 M imidazole gradient. To remove fragments derived from partial proteolytic degradation, the eluted material was purified by RP-HPLC with a Zorbax 300SB-CN column (9.4 \times 250 mm, 5 μ m, Agilent) using a 0–40% gradient of 0.1% TFA in H₂O and 0.1% TFA in CH₃CN and freeze-dried. The N-terminal His₆-tag was removed using a recombinant dipeptidyl aminopeptidase I (DAPase) containing a C-terminal His-tag (TAGzyme, Qiagen) for 2 h at 37 °C 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 His₂ dipeptides. The protein was subjected to a final RP-HPLC step and analyzed by LC-MS on a Gilson HPLC system coupled with an ESI-MS single quadrupole mass spectrometer (Applied Biosystems API-150EX), using a Zorbax 300SB-CN column (2.1 \times 150 mm, 5 μ m, Agilent) and a 0–50% gradient of 0.1% TFA in H₂O and 0.1% TFA in CH₃CN. 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 (M_r calculated: 14894 Da; M_r observed: 14893 Da). The purified protein was freeze-dried and used for spectroscopic studies. The yield was \sim 8 mg protein/1 L of culture.

2. Purification in Native Conditions. 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 \sim 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, and 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, and 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 and 5 mM DTT, pH 7.4). The His₆-DLL4_IC protein was purified by ion-exchange chromatography on an 8 \times 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, and 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.

3. ¹⁵N Isotopic Enrichment. One liter of ¹⁵N-M9 minimal medium (6 g/L Na₂HPO₄, 3 g/L KH₂PO₄, 0.5 g/L NaCl, 1.3 g/L glycerol, 0.5 g/L ¹⁵NH₄Cl, 0.12 g/L MgSO₄, and 0.01 g/L CaCl₂) containing nutrients supplemented as yeast nitrogen base without amino acids or NH₄SO₄ (1.7 g/L), pH 7, with 100 μ g/mL ampicillin and 25 μ g/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 \sim 0.6 and protein expression was induced with IPTG (1 mM) for 4.5 h at 37 °C. Protein purification was carried out in denaturing conditions as described above, with a final yield of \sim 8 mg of pure product/1 L of culture.

4. Expression of Δ N-DLL4_IC and Purification from Inclusion Bodies. One liter of LB containing 100 μ g/mL ampicillin and 25 μ g/mL chloramphenicol was inoculated with a clone of BL21(DE3)pLysS cells transformed with the Δ N-dll4_IC construct. Cells were grown at 37 °C to an OD of \sim 0.9 and protein expression was 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, MgSO₄ (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 and 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 \times 250 mm, 5 μ m, Agilent) using a 0–60% gradient of 0.1% TFA in H₂O and 0.1% TFA in CH₃CN. The freeze-dried product was analyzed by LC-MS which confirmed the correct molecular

size (M_r calculated: 1267.9 Da; M_r observed: 1265.0 Da). The yield was ~ 4 mg protein ($\sim 95\%$ purity) /1 L of culture.

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 semipreparative RP-HPLC on a Zorbax 300SB-C18 column (9.4×250 mm, $5 \mu m$, 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 were 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 and 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 ($R_{S,N}$) and fully unfolded ($R_{S,U}$) proteins of known molecular weight (MW) were determined according to the equations:³⁷ $\log(R_{S,N}) = -(0.254 \pm 0.002) + (0.369 \pm 0.001) \log(MW)$, and $\log(R_{S,U}) = -(0.543 \pm 0.004) + (0.502 \pm 0.001) \log(MW)$.

Circular Dichroism (CD). The freeze-dried protein powder was dissolved either in 5 mM MES buffer and 1 mM TCEP, pH 6.3, or in 5 mM Tris buffer and 1 mM TCEP, pH 7.5. Protein concentration was determined by UV absorbance at 280 nm using the calculated ϵ value of $2560 \text{ M}^{-1} \text{ cm}^{-1}$. CD spectra of solutions of DLL4_IC, ΔN -DLL4_IC 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 Dichroweb³⁸ tools CDSSTR, CONTIN and SELCON. The helical content was also determined from the mean residue ellipticity (MRE, $(\text{deg} \cdot \text{cm}^2)/\text{dmol}$) at 222 nm ($[\theta]_{222}$) according to the equation: $[\alpha] = 100 \cdot [\theta]_{222}/\theta_f$ and $\theta_f = -40\,000 \cdot (1 - 2.57/n)$ 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.³⁹

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

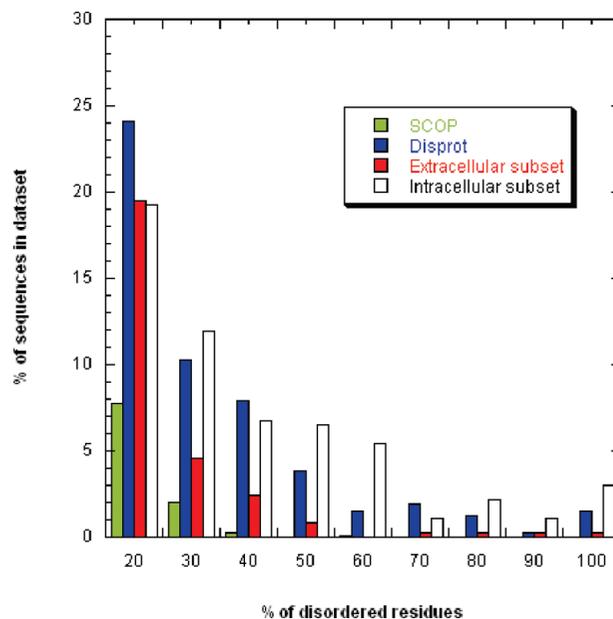


Figure 2. Disorder predictions. Percentage of sequences in data sets (reduced SCOP, DisProt, extracellular and intracellular regions of type I single-pass receptors) containing $\geq 20\%$ predicted disordered residues. Intrinsically disordered regions were computed by DisEMBL using the Remark465 definition.

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 DIPS mixing time. Data were transformed using X-WinNMR (Bruker) and analyzed using CARA (<http://www.nmr.ch>). 1H chemical shifts were referenced to internal DSS.

Results

Intrinsic Disorder Is Prevalent in the Cytoplasmic Region of Human Single-Pass Transmembrane Receptors. A data set 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,²² the database of protein disorder (469 sequences), and a reduced set (1357 sequences) extracted from SCOP,²³ a database of domains of known structure, were used as control data sets. ID regions in each of these data sets were predicted by subjecting the sequences to DisEMBL.²⁴ Figure 2 shows the fraction of sequences in the data sets versus the fraction ($\geq 20\%$) of residues predicted to be disordered. First, a surprisingly high incidence of ID is observed in the intracellular subset, as it contains a significantly higher fraction of disordered residues relatively to the DisProt data set. 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, respectively²⁴), whereas in the latter, the incidence of ID is only slightly higher than that computed for the SCOP domain database. IUPred predictions are consistent with

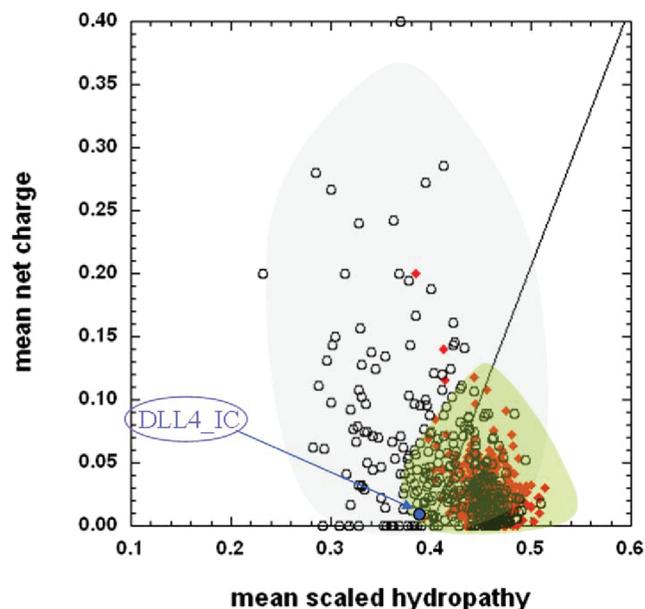


Figure 3. Intrinsic disorder. The absolute value of the mean net charge is plotted versus the mean scaled Kyte-Doolittle hydropathy for the intracellular subset (circles) and extracellular subset (diamonds). The shaded areas contain 90% of the points from the data set of disordered regions (DisProt, gray) and the data set of domains of known structure (SCOP, green). The border between structured and natively disordered proteins is drawn as a line.

DisEMBL results. The fraction of disordered residues is higher in intracellular regions, as calculated from pairwise energies (see Supporting Information). Overall, the percentage 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.

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.²⁶ Consequently, intrinsically disordered proteins generally localize within a unique region of the charge/hydropathy phase space, whereas ordered proteins cluster in a separate region (gray and green shaded areas in Figure 3, respectively).⁴⁰ In Figure 3, 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 averaged 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.

The amino acid compositional analysis confirms these observations. Figure 4 shows a comparison of the amino acid compositions of the intra- and extracellular subsets along with the comparison between the two control data sets (ordered and disordered proteins). With few exceptions (see Discussion), the intracellular set is depleted in the order-promoting residues⁴¹ (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 data sets of ordered and disordered proteins.

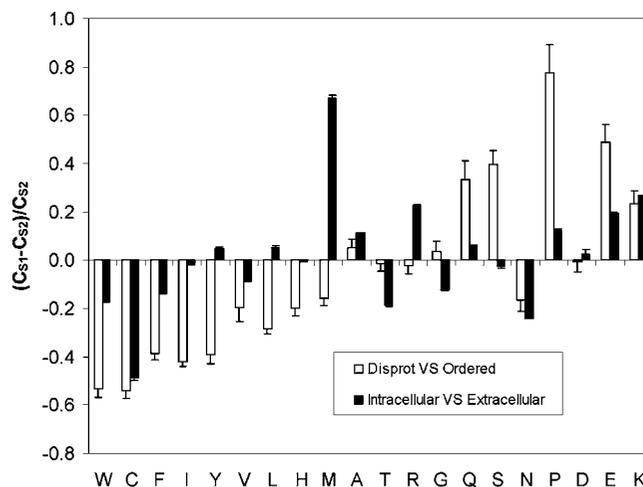


Figure 4. Amino acid compositional analysis. Enrichment or depletion in each amino acid type appears as a positive or negative bar, respectively. Amino acids are indicated by the single-letter code and ordered according to increasing flexibility. Order-promoting residues, W, C, F, I, Y, V, L, N; disorder-promoting 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 ID in human single-pass transmembrane receptors, and that ID is highly concentrated in the cytoplasmic region.

A Case Study: the Cytoplasmic Region of the Notch Ligand Delta-4. 1. Sequence Analysis.

The amino acid sequence (Figure 5a) of the cytoplasmic region of Notch ligand Delta-4 (DLL4_IC) was retrieved from the intracellular subset (blue circle in the mean net charge/hydropathy plot in Figure 3) and subjected to disorder and secondary structure predictions. PONDR,³¹ a predictor-based neural network trained with sequences of intrinsically disordered regions, predicted four disordered stretches that account for 37% of the entire sequence of DLL4_IC (Figure 5b), whereas DisEMBL predicted 10% (Remark465) and 55% (Hot Loop) of disordered residues. DLL4_IC is expected to adopt some secondary structure, as suggested by different secondary structure predictors (PSIPRED,³² JNet,³³ SSpro³⁴). 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 5a).

2. Protein Expression and Purification. The recombinant protein corresponding to the intracellular region of human Delta-like 4 (DLL4_IC, residues 553–685 of DLL4_HUMAN, 133 amino acids) (Figure 5a) was expressed in *E. coli* from a synthetic gene designed to optimize the codon usage for heterologous expression (see Supporting Information). Despite an extensive proteolytic degradation, the final material was highly pure (>95%, as determined by RP-HPLC), and could be recovered with a good yield (8 mg/L), thus, allowing its characterization by circular dichroism, NMR and size exclusion chromatography. The truncated protein Δ N-DLL4_IC was expressed in a similar system, and recovery from inclusion bodies allowed a single-step purification by RP-HPLC.

3. DLL4_IC Is Mainly Disordered in Solution. The presence of secondary structure in DLL4_IC was investigated by circular dichroism spectroscopy (CD). The far-UV CD spectrum of DLL4_IC (Figure 6a) in Tris buffer shows a strong minimum at

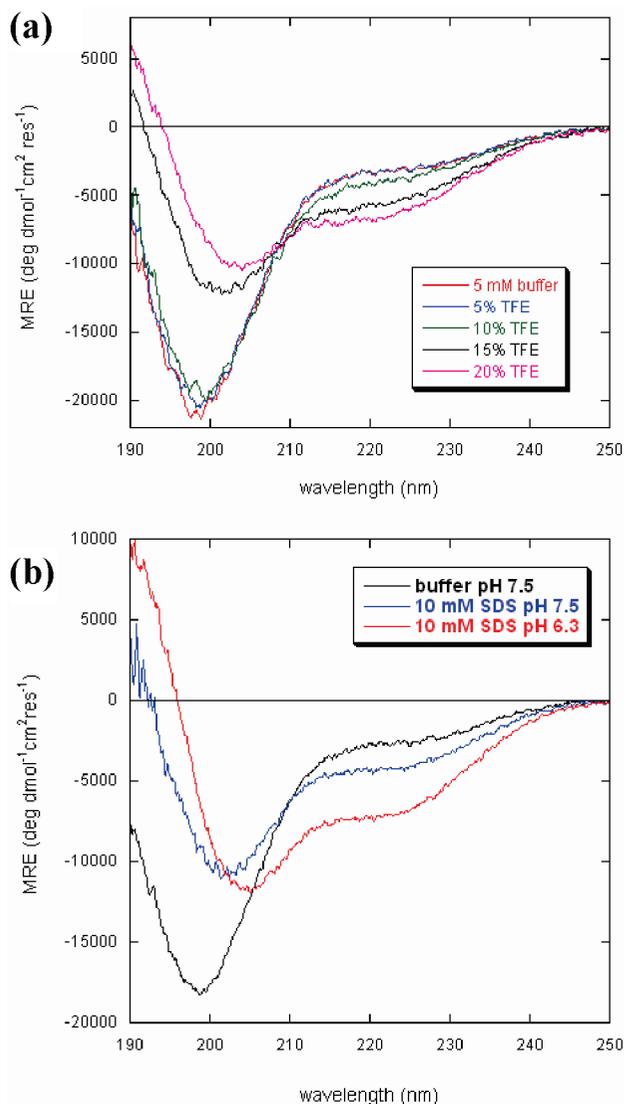


Figure 6. Circular dichroism. Far-UV CD spectra of DLL4_IC (7.6 μ M) in 5 mM Tris-HCl buffer and 1 mM TCEP, pH 7.5, and in the presence of increasing concentrations of TFE (5, 10, 15, 20%, v/v) (a) or 10 mM SDS at pH 7.5 or 6 (b).

change in the secondary structure was observed upon the 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 6a). The helical content increased from 3% to 17% upon TFE addition (0–20%, v/v), with a drastic change in ellipticity already observed between 10% and 15% TFE. Also, a significant increase in β -strand and turns structure was 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 toward the α -helical structure, reaching a maximum of \sim 8% of α -helix at saturation (10 mM SDS, see Supporting Information). Interestingly, at the same saturating SDS concentration (10 mM), the α -helical content experiences a significant increase as the pH decreases (8% of α -helix at pH 7.5 versus 23% at pH 6.3, Figure 6b), while the same shift in pH only slightly increases the β -strand content of the protein in buffer.

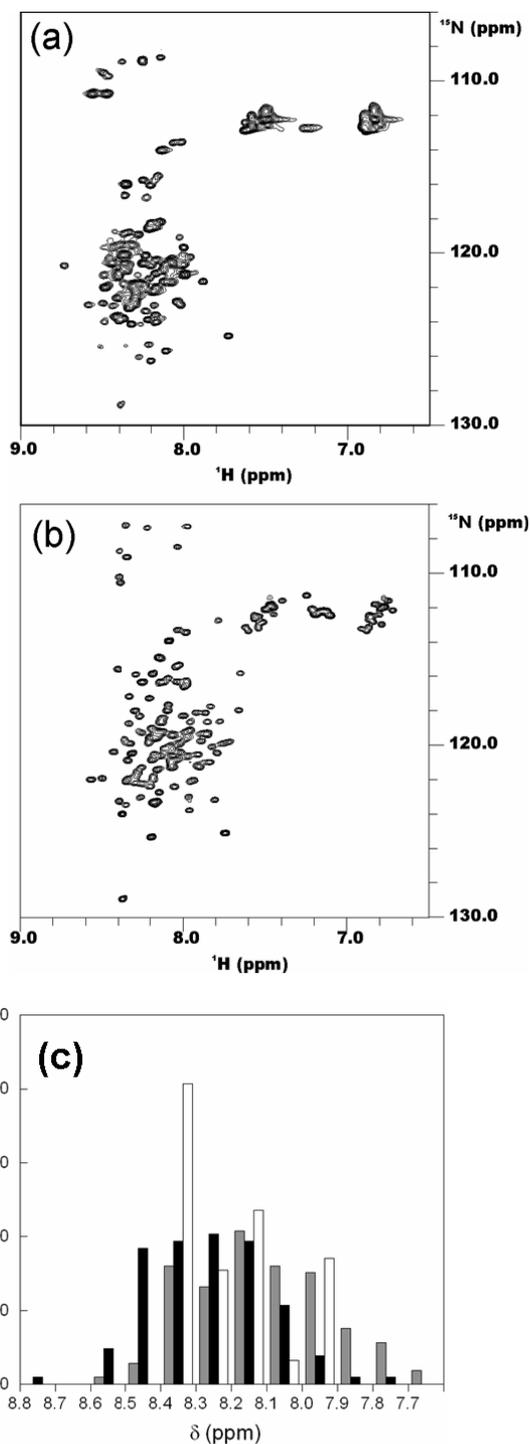


Figure 7. NMR. ^1H – ^{15}N HSQC spectra of DLL4_IC (0.5 mM) in $\text{H}_2\text{O}/\text{D}_2\text{O}$ (90/10, v/v), 4 mM TCEP, pH 5.6, recorded at 303 K in the absence (a) or presence (b) of SDS (50 mM); (c) distribution of ^1HN chemical shifts of DLL4_IC in $\text{H}_2\text{O}/\text{D}_2\text{O}$ (black bars), in the presence of SDS (gray bars) and for random coil values for a protein of the same amino acid composition (white bars).

The conformation of DLL4_IC in the presence of SDS micelles was further analyzed by NMR. The ^1H – ^{15}N HSQC spectrum of DLL4_IC obtained at saturating concentrations of SDS is somewhat different from that of the protein alone (Figure 7a,b). 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

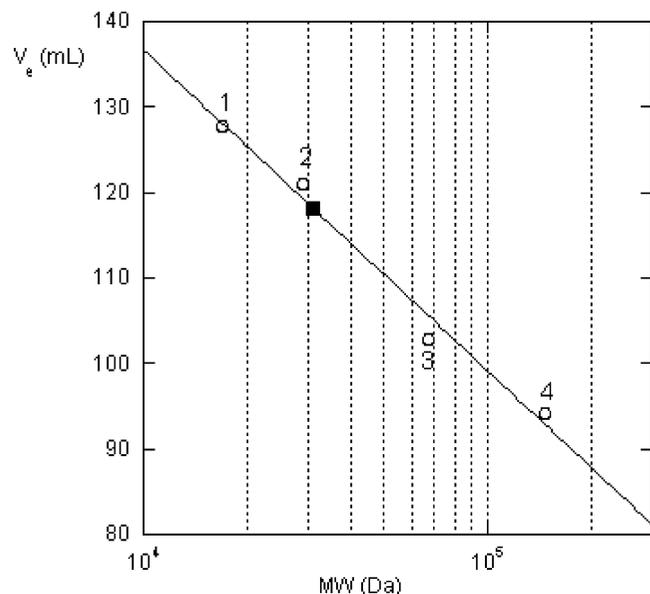


Figure 8. Size exclusion chromatography. Calibration standards are shown as open circles (1, lactate dehydrogenase [147 kDa]; 2, bovine serum albumin [67 kDa]; 3, carbonic anhydrase [29 kDa]; 4, horse myoglobin [17 kDa]), DLL4_IC as a filled square (apparent MW = 31 kDa). The calibration curve ($R = 0.99$) is also shown. V_e = elution volume.

resonances are still clustered in a relatively narrow region (7.7–8.4 ppm), but the average value of ^1HN chemical shifts (8.11 ppm) is smaller and the dispersion slightly larger ($\sigma = 0.20$) compared to the values obtained for the protein alone (Figure 7c). Moreover, ~ 90 cross-peaks could be counted in the HN-H α region of the ^1H - ^{15}N HSQC-TOCSY spectrum (see Supporting Information), most of them in the 4.2–4.6 ppm region. Significantly, a discrete number of nonglycine HN-H α cross-peaks displayed a high field shift (< 4.2 ppm). The lack of a significant chemical shift dispersion in the HN and H α chemical shifts, even in the presence of SDS micelles, is evidence for the 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 $^1\text{H}\alpha$, $^{13}\text{C}\alpha$, and $^{13}\text{C}'$ have been widely used to map regions with well-defined secondary structure, but require residue-specific sequential assignments of the backbone resonances. Recently, it has been shown that backbone ^1HN and ^{15}NH chemical shifts are also somewhat sensitive to secondary structure.⁴⁴ The small upfield shift of these and of selected H α resonances in the presence of SDS might then be explained in terms of partial α -helical formation, consistent with CD results. It cannot be ruled out, however, that the negatively charged headgroup of SDS can also contribute to the upfield shift. It must be noted 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).

5. The C-Terminal Region of DLL4_IC Is Plastic. To locate the region(s) of DLL4_IC that adopt the observed secondary structures based on the environment, we broke down the DLL4_IC sequence into four fragments (ΔN -DLL4_IC and peptides P1, P2, and P3), as indicated in Figure 5a, and studied their conformation alone or in the presence of TFE or SDS using

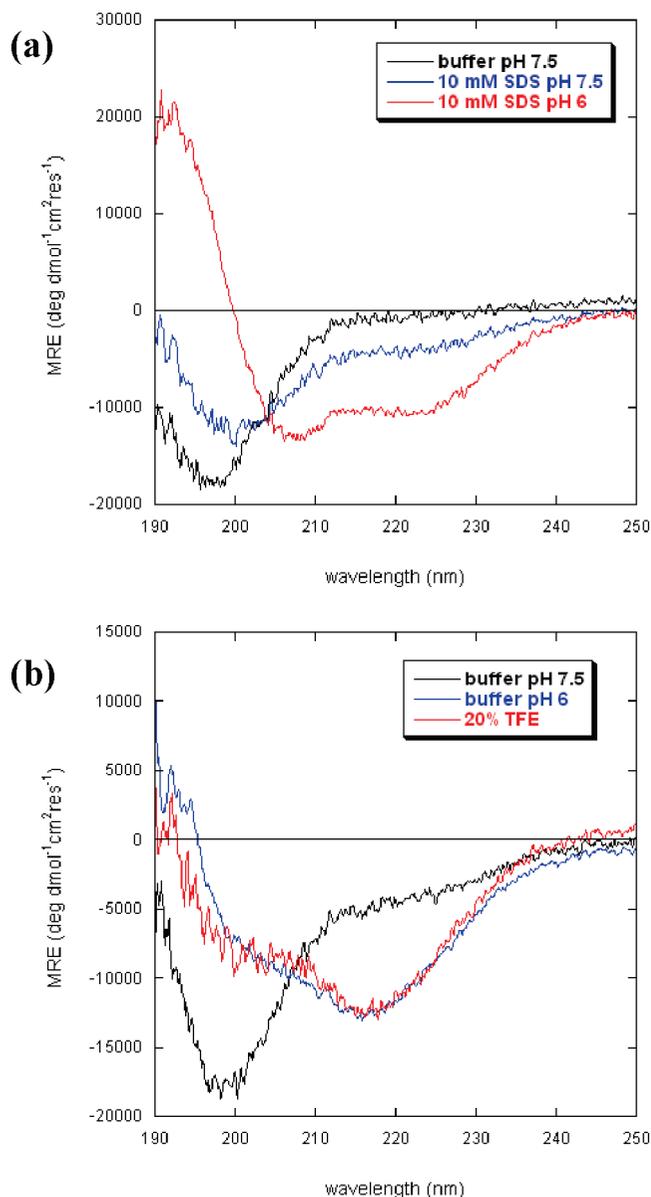


Figure 9. Circular dichroism. Far-UV CD spectra of P3 (17.9 μM) in (a) 5 mM Tris-HCl buffer, 1 mM TCEP, in the presence of 10 mM SDS at pH 7.5 or 6.0, and in 20% TFE or in (b) 5 mM Tris-HCl buffer (pH 7.5 or 6) or 20% TFE.

far-UV CD spectroscopy. The DLL4_IC sequence was split at proline sites, because of its helix-breaking properties, and boundaries were selected based on the predictions by PONDR (Figure 5b). The procedure for generating the fragments is described in the Material and Methods. Our results show that the central region covered by P1 and P2 is disordered and does not adopt any secondary structure in all tested conditions (Supporting Information). However, the region encompassing the last 24 residues is highly plastic (P3, Figure 9, and Supporting Information) and, starting from a prevalent coil conformation in buffer at neutral pH, it can adopt (i) a mainly strand conformation in the presence of TFE or when the pH is slightly acidic (appearance of the negative band at 219 nm, typical of β -strand structures) or (ii) a α -helical conformation in the presence of SDS micelles when the pH is dropped from neutrality to 6. All these conformational changes are fully reversible (data not shown). Overall, these data indicate that the last 24 C-terminal residues account for most of the plasticity

Intrinsic Disorder in the Delta-4 Cytoplasmic Region

displayed by DLL4_IC and that the pH-dependent conformational switch in the full-length DLL4_IC, in the presence of SDS micelles, is due to the formation of an ~ 10 residue α -helix in the C-terminal region, as assessed from the changes in the mean residue ellipticity at 222 nm. The possible implications of these findings are discussed below.

Discussion

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 data set (369) by the SMART tool¹⁴ revealed that a 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 (Figure 3), 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 a 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.⁴⁵

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 this uncharacterized subset and were experimentally proven to be intrinsically disordered.⁴⁶ Previously, we showed that the cytoplasmic tails of the five different human Notch ligands display little sequence similarity between one another, display no homology with sequences of known fold, and are predicted to be disordered.⁴⁷ We have also shown that a recombinant protein corresponding to the intracellular region of the Notch ligand Jagged-1 is actually disordered in solution.⁴⁸ Here, for the first time, we prove, through a combination of predictions and experimental biophysical methods, that a fragment spanning the intracellular region of human Delta-4 is intrinsically disordered in solution, yet at the same time, in response to changes in the physicochemical environment, can form interconvertible secondary structures through its plastic C-terminal region (P3). Delta-4 was shown to interact through its evolutionarily conserved C-terminal motif (ATEV) (Figure 5c) with the PDZ domains of Dlg-1, the human homologue of *Drosophila* Discs Large protein.¹⁶ In the available crystal structures, peptides bearing a PDZ recognition motif bind in a groove between β B and α B on the surface of the PDZ domain and adopt a β -strand conformation, stitched as an additional strand into the anti-parallel β -sheet on the surface of the PDZ domain.⁴⁹ The mainly β structure seen in P3 in the presence of TFE or at slightly acidic pH, may be representative of the conformation that the C-terminal region of Delta-4 adopts when bound to Dlg-1. We are currently exploring the possibility that the C-terminal region of Delta-4 is preorganized in a β -hairpin conformation, ready to dock the target PDZ domain. A stable β -hairpin at the C-terminus, or a significant population in the same conformation, would drastically reduce the entropic cost of binding to the target PDZ domain. As the C-terminal region is very well conserved through species (Figure 5c), it is likely that this preorganization is not limited to human Delta-4, but applies to all the other homologues. The coexistence of global disorder (i.e., lack of a well-defined globular structure) and local preorganization (i.e., the propensity to form certain types of secondary structures locally, either in a stable or transient way) may indeed represent a general mechanism to increase binding specificity while maintaining structural flexibility. Such regions might be identified through a combination of secondary structure and disorder prediction tools and assayed experimentally.

The mainly helical conformation of P3 in the presence of SDS may, on the other hand, be representative of the membrane-bound, uncomplexed form of Delta-4. 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.⁵⁰ 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 homogeneous, but decreases to ~6.0 in proximity of the membrane.⁵¹ 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.

Alternatively, the observed secondary structures might be induced through the binding of Delta-4'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 Delta-1 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.²⁰

Overall, there is convergent evidence that the asymmetric disorder distribution observed in Delta-4 and its homologues reflects a more general phenomenon in transmembrane receptors of the same class, where the ordered extracellular domains appear to act as rigid scaffolds for ligand binding, while the flexible intracellular tails transduce the signal within the cell and activate the complex cellular response. Consequently, we speculate that the intracellular tails carry out their function by binding possibly multiple partners, either remaining unstructured, by exploiting preorganized 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.

Abbreviations: ID, intrinsic disorder; DLL4_IC, human Delta-like 4 cytoplasmic region; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate-*d*₆ sodium salt; DTT, DL-dithiothreitol; GuHCl, guanidine hydrochloride; HSQC, heteronuclear single quantum correlation; IMAC, immobilized metal ion affinity chromatography; IPTG, isopropyl β -D-1-thiogalactopyranoside; LC-MS, liquid chromatography coupled with mass spectrometry; PDZ, domain present in PSD-95, Dlg, and ZO-1/2; RP-HPLC, reverse-phase high-performance liquid chromatography; TCEP, Tris(2-carboxyethyl)phosphine hydrochloride; TFA, trifluoroacetic acid; TFE, 2,2,2-trifluoroethanol.

Acknowledgment. We thank Dorian Lamba (CNR-ELETTRA, Trieste, Italy) for use of the CD spectropolarimeter. We are grateful to Fabio Calogiuri (CERM, Sesto Fiorentino, Italy) for technical assistance with the acquisition of NMR spectra. We acknowledge the support of the EU (European Network of Research Infrastructures for Providing Access and Technological Advancements in Bio-NMR) for access to the CERM NMR facility. The support of the IUPUI Signature Centers Initiative is gratefully acknowledged. We also thank Zsuzsanna Dosztányi (Institute of Enzymology, Budapest, Hungary) for the IUPred code, Mircea Pacurar (ICGEB) and Somdutta Dhir (ICGEB) for writing the scripts used in disorder analysis, Ruth W. Craig (Dartmouth Medical School) for critical reading of the manuscript, and Maristella Cogliervina (ICGEB) for her useful suggestions. This work is part of A. De Biasio's Ph.D. thesis.

Supporting Information Available: Figure S1, codon usage optimization. Figure S2, (a) gene synthesis, agarose gel (1%) of the gene assembly PCR mixture; (b) protein purification, Coomassie Blue-stained SDS-PAGE (4–12%) of DLL4_IC before

and after His₆-tag removal. Figure S3, purification in native conditions. RP-HPLC analysis of His₆-DLL4_IC purified by IMAC followed by ion exchange chromatography. Figure S4, far-UV CD spectrum of His₆-DLL4_IC (14.1 μ M) in 5 mM phosphate buffer, 1 mM TCEP, pH 7.5, purified in native conditions. Figure S5, size exclusion chromatography. Elution profile of DLL4_IC on a Sephacryl S-200 column. Figure S6, NMR. ¹H–¹⁵N HSQC-TOCSY of DLL4_IC in H₂O/D₂O (90/10 v/v) containing 4 mM TCEP, 2 mM EDTA-d₁₆, 15 μ M DSS, pH 5.6, protein concentration ~0.5 mM. Figure S7, NMR. 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 μ M DSS, pH 5.6, protein concentration ~0.5 mM. Figure S8, CD in the presence of SDS. (a) Far-UV CD spectra of DLL4_IC (7.6 μ M) in 5 mM Tris-HCl buffer, 1 mM TCEP, pH 7.5, in the presence of different concentrations of SDS (mM); (b) DLL4_IC helix-enrichment upon addition of increasing amounts of SDS (pH 7.5). Figure S9, CD. Far-UV CD spectra of (a) DLL4_IC (7.6 μ M), (b) Δ N-DLL4_IC (10.2 μ M), (c) P1 (14 μ M), (d) P2 (18.7 μ M) and (e) P3 (17.9 μ M) 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. Figure S10, 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. Sequences of the extra- and intracellular regions of human single-pass transmembrane receptors with the type I topology used in this study are available on request. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Tompa, P.; Szasz, C.; Buday, L. Structural disorder throws new light on moonlighting. *Trends Biochem. Sci.* **2005**, *30*, 484–489.
- (2) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell. Biol.* **2005**, *6*, 197–208.
- (3) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **2004**, *337*, 635–645.
- (4) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **2002**, *323*, 573–584.
- (5) Mohan, A.; Oldfield, C. J.; Radivojac, P.; Vacic, V.; Cortese, M. S.; Dunker, A. K.; Uversky, V. N. Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.* **2006**, *362*, 1043–1059.
- (6) Minezaki, Y.; Homma, K.; Kinjo, A. R.; Nishikawa, K. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.* **2006**, *359*, 1137–1149.
- (7) Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. Intrinsic disorder in transcription factors. *Biochemistry* **2006**, *45*, 6873–6888.
- (8) Minezaki, Y.; Homma, K.; Nishikawa, K. Intrinsically disordered regions of human plasma membrane proteins preferentially occur in the cytoplasmic segment. *J. Mol. Biol.* **2007**, *368*, 902–913.
- (9) Hurlbut, G. D.; Kankel, M. W.; Lake, R. J.; Artavanis-Tsakonas, S. Crossing paths with Notch in the hyper-network. *Curr. Opin. Cell Biol.* **2007**, *19*, 166–175.
- (10) Bray, S. J. Notch signalling: a simple pathway becomes complex. *Nat. Rev. Mol. Cell. Biol.* **2006**, *7*, 678–689.
- (11) Artavanis-Tsakonas, S.; Rand, M. D.; Lake, R. J. Notch signaling: cell fate control and signal integration in development. *Science* **1999**, *284*, 770–776.
- (12) Noguera-Troise, I.; Daly, C.; Papadopoulos, N. J.; Coetzee, S.; Boland, P.; Gale, N. W.; Lin, H. C.; Yancopoulos, G. D.; Thurston, G. Blockade of Dll4 inhibits tumour growth by promoting non-productive angiogenesis. *Nature* **2006**, *444*, 1032–1037.
- (13) Ridgway, J.; Zhang, G.; Wu, Y.; Stawicki, S.; Liang, W. C.; Chantry, Y.; Kowalski, J.; Watts, R. J.; Callahan, C.; Kasman, I.; Singh, M.; Chien, M.; Tan, C.; Hongo, J. A.; de Sauvage, F.; Plowman, G.; Yan,

- M. Inhibition of Dll4 signalling inhibits tumour growth by deregulating angiogenesis. *Nature* **2006**, *444*, 1083–1087.
- (14) Letunic, I.; Copley, R. R.; Pils, B.; Pinkert, S.; Schultz, J.; Bork, P. SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res.* **2006**, *34*, D257–260.
- (15) Le Borgne, R.; Bardin, A.; Schweisguth, F. The roles of receptor and ligand endocytosis in regulating Notch signaling. *Development* **2005**, *132*, 1751–1762.
- (16) Six, E. M.; Ndiaye, D.; Sauer, G.; Laabi, Y.; Athman, R.; Cumano, A.; Brou, C.; Israel, A.; Logeat, F. The notch ligand Delta1 recruits Dlg1 at cell-cell contacts and regulates cell migration. *J. Biol. Chem.* **2004**, *279*, 55818–55826.
- (17) LaVoie, M. J.; Selkoe, D. J. 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*, 34427–34437.
- (18) Six, E. M.; Ndiaye, D.; Laabi, Y.; Brou, C.; Gupta-Rossi, N.; Israel, A.; Logeat, F. The Notch ligand Delta1 is sequentially cleaved by an ADAM protease and gamma-secretase. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 7638–7643.
- (19) Ikeuchi, T.; Sisodia, S. S. The Notch ligands, Delta1 and Jagged2, are substrates for presenilin-dependent “gamma-secretase” cleavage. *J. Biol. Chem.* **2003**, *278*, 7751–7754.
- (20) Hiratochi, M.; Nagase, H.; Kuramochi, Y.; Koh, C. S.; Ohkawara, T.; Nakayama, K. 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*, 912–922.
- (21) Ascano, J. M.; Beverly, L. J.; Capobianco, A. J. The C-terminal PDZ-ligand of JAGGED1 is essential for cellular transformation. *J. Biol. Chem.* **2003**, *278*, 8771–8779.
- (22) Sickmeier, M.; Hamilton, J. A.; LeGall, T.; Vacic, V.; Cortese, M. S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V. N.; Obradovic, Z.; Dunker, A. K. DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* **2007**, *35*, D786–793.
- (23) Andreeva, A.; Howorth, D.; Chandonia, J. M.; Brenner, S. E.; Hubbard, T. J.; Chothia, C.; Murzin, A. G. Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.* **2007**.
- (24) Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B. Protein disorder prediction: implications for structural proteomics. *Structure* **2003**, *11*, 1453–1459.
- (25) Dosztanyi, Z.; Csizmok, 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*, 827–839.
- (26) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are “natively unfolded” proteins unstructured under physiologic conditions. *Proteins* **2000**, *41*, 415–427.
- (27) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. Sequence complexity of disordered protein. *Proteins* **2001**, *42*, 38–48.
- (28) Vacic, V.; Uversky, V. N.; Dunker, A. K.; Lonardi, S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinf.* **2007**, *8*, 211.
- (29) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (30) Vihinen, M.; Torkkila, E.; Riikonen, P. Accuracy of protein flexibility predictions. *Proteins* **1994**, *19*, 141–149.
- (31) Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K. Predicting intrinsic disorder from amino acid sequence. *Proteins* **2003**, *53*, 566–572.
- (32) Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **1999**, *292*, 195–202.
- (33) Cuff, J. A.; Barton, G. J. Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins* **1999**, *34*, 508–519.
- (34) 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*, 228–235.
- (35) Hoover, D. M.; Lubkowski, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **2002**, *30*, e43.
- (36) Medigue, C.; Rouxel, T.; Vigier, P.; Henaut, A.; Danchin, A. Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J. Mol. Biol.* **1991**, *222*, 851–856.
- (37) Uversky, V. N. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry* **1993**, *32*, 13288–13298.
- (38) Lobley, A.; Whitmore, L.; Wallace, B. A. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* **2002**, *18*, 211–212.
- (39) Chen, Y. H.; Yang, J. T.; Chau, K. H. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* **1974**, *13*, 3350–3359.
- (40) Uversky, V. N. What does it mean to be natively unfolded? *Eur. J. Biochem.* **2002**, *269*, 2–12.
- (41) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. Intrinsically disordered protein. *J. Mol. Graphics Modell.* **2001**, *19*, 26–59.
- (42) Wishart, D. S.; Sykes, B. D.; Richards, F. M. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **1991**, *222*, 311–333.
- (43) Roccatano, D.; Colombo, G.; Fioroni, M.; Mark, A. E. 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*, 12179–12184.
- (44) Wang, Y.; Jardetzky, O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.* **2002**, *11*, 852–861.
- (45) von Heijne, G. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **1989**, *341*, 456–458.
- (46) Sigalov, A. B.; Aivazian, D. A.; Uversky, V. N.; Stern, L. J. Lipid-binding activity of intrinsically unstructured cytoplasmic domains of multichain immune recognition receptor signaling subunits. *Biochemistry* **2006**, *45*, 15731–15739.
- (47) 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.
- (48) 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*, 5325–5336.
- (49) Jelen, F.; Oleksy, A.; Smietana, K.; Otlewski, J. PDZ domains—common players in the cell signaling. *Acta Biochim. Pol.* **2003**, *50*, 985–1017.
- (50) Olivotto, M.; Arcangeli, A.; Carla, M.; Wanke, E. Electric fields at the plasma membrane level: a neglected element in the mechanisms of cell signalling. *BioEssays* **1996**, *18*, 495–504.
- (51) Slavik, J. Intracellular pH topography: determination by a fluorescent probe. *FEBS Lett.* **1983**, *156*, 227–230.

PR800063U