Structural and dynamic characterization of the C-terminal tail of ErbB2: disordered but not random

L. Pinet, Y.-H. Wang, C. Deville, E. Lescop, F. Guerlesquin, A. Badache, F. Bontems, Nelly Morellet, N. Assrir, C. van Heijenoort, et al.

Manuscript submitted to **Biophysical** *Journal*
Article

Structural and dynamic characterization of the C-terminal tail of ErbB2: disordered but not random

L. Pinet[1,2], Y.-H. Wang[1,3], C. Deville[1,4], E. Lescop[1], F. Guerlesquin[5], A. Badache[6], F. Bontems[1,7], N. Morellet[1], D. Durand[8], N. Assrir[1], and C. van Heijenoort[1,*]

1
2
3
4
5
6
7
8

[*]Correspondence: @

ABSTRACT ErbB2 (or HER2) is a receptor tyrosine kinase overexpressed in some breast cancers, associated with poor prognosis. Treatments targeting the receptor extracellular and kinase domains have greatly improved disease outcome in the last twenty years. In parallel, the structures of these domains have been described, enabling better mechanistic understanding of the receptor function and targeted inhibition. However, ErbB2 disordered C-terminal cytoplasmic tail (CtErbB2) remains very poorly characterized in terms of structure, dynamics and detailed functional mechanism. Yet, it is where signal transduction is triggered, via phosphorylation of tyrosine residues, and carried out, via interaction with adaptor proteins. Here we report the first description of ErbB2 disordered tail at atomic resolution, using NMR and SAXS. We show that although no part of CtErbB2 has any stable secondary or tertiary structure, it has around 20% propensity for a N-terminal helix that is suspected to interact with the kinase domain, and many PPII stretches distributed all along the sequence, forming potential SH3 and WW domains binding sites. Moreover, we identified a long-range transient contact involving CtErbB2 termini. These characteristics suggest new potential mechanisms of auto-regulation and protein-protein interaction.

SIGNIFICANCE We report here the first description of the receptor tyrosine kinase ErbB2 disordered tail (CtErbB2) at atomic resolution, using NMR and SAXS. We show that although CtErbB2 exhibits no stable structure, it does exhibit partial secondary and tertiary structures likely important for its function. These structural elements are consistent with an active role of the C-terminal tail in the regulation of the receptor's activity, thanks to the presence of preformed structures for intramolecular interactions, as well as long-range contacts modulating accessibility of those sites and proline interaction sites distinct from the main tyrosine sites. Together, those results reinforce the view that disordered tails of receptors are more than random anchors for partners.

INTRODUCTION

The ErbB proteins (ErbB1/EGFR/HER1, ErbB2/HER2/neu, ErbB3/HER3 and ErbB4/HER4) are receptor tyrosine kinases that have been extensively studied following the discovery of their involvement in different types of cancer (1, 2). They are constituted of an extracellular domain involved in ligand binding and receptor dimerization, a transmembrane helix, and an intracellular part made of a tyrosine kinase domain and a C-terminal tail. So far, major steps of ErbB signaling have been identified: ligand binding to ErbB1, ErbB3 or ErbB4 induces a conformational rearrangement of their extracellular domain from a tethered into an "open" form that is poised for homo- or heterodimerization. The kinase domains are activated in the dimers (or potential higher-order oligomers (3)), leading to tyrosine phosphorylation of the C-terminal tails (4). This tail is a hub for phosphorylation-regulated protein-protein interactions with adaptor proteins that trigger downstream signaling, including the RAS/MAPK, PI3K/Akt, Src kinases and STAT transcription factors dependent pathways (4).

ErbB2 has no known ligand, but has a constitutively "open" conformation of its extracellular domain, making it the preferential dimerization partner for other ErbB receptors (5). ErbB2 upregulation is found in up to 20% of breast cancers and correlates with poor prognosis (6), and therefore sparked a lot of efforts for the development of targeting strategies, including monoclonal antibodies binding the extracellular domain and small molecules inhibiting its tyrosine kinase activity (7, 8). These strategies have been widely successful, but are limited by inherent or acquired resistance, calling for new combined therapeutic approaches and dual HER2-targeting (9) that require an extended knowledge of associated molecular mechanisms. While elucidating the structures of the extracellular (10–18), transmembrane (19–21), juxtamembrane (21) and kinase (22, 23) domains has given extensive mechanistic insights, ErbB2 C-terminal tail (CtErbB2) is the only region for which we lack such structural description. It was shown that deletion of this region completely abolishes the transforming potential of the activated receptor, while mutation of its five autophosphorylation sites to phenylalanine reduces it by 92% (24). Each phospho-tyrosine interacts with distinct partners, SH2 and PTB domains-containing proteins or MEMO (4, 25, 26), to trigger different signaling pathways *in vivo* (24, 27). However, the structural features underlying these properties are poorly described. Only a few studies on the C-terminal tail of EGFR, the most studied ErbB receptor, were conducted, and it was suggested that its conformation is dependent on its phosphorylation state and regulates kinase activity (28, 29). Similar mechanisms could be at stake in CtErbB2 function.

We previously performed the NMR assignment of CtErbB2 and demonstrated that it is intrinsically disordered (30), in agreement with predictions based on its sequence composition and with measurements of solvent accessibility (31). Intrinsically disordered proteins (IDPs) are a class of proteins with no major, stable tridimensional structure. Instead, their conformation can be described by an ensemble of interconverting flexible structures. IDPs are over-represented in cell-cycle control and signal transduction (32), most likely due to their specific mechanical properties, their increased capture radius ("fly-casting" mechanism (33, 34)) and their specific but usually low- to medium-affinity binding (35). This binding generally involves short linear motifs (SLiMs) and molecular recognition features (MoRFs) that often fold upon binding, and regulate accessibility through long-range contacts that can be modulated by post-translational modifications. Although the development of drugs inhibiting interactions that involve IDPs is challenging, some strategies are beginning to emerge (36). The description of the features of CtErbB2 unphosphorylated apo-state would be the first step towards a better understanding of the mechanisms that lead to its phosphorylation and interaction with adaptor proteins.

Solution state NMR spectroscopy is the dedicated atomic resolution method for such investigation of conformational heterogeneity and dynamics, and is used extensively for determination of conformational ensembles, combined with small angle X-ray scattering (SAXS) (37). Here we present the first atomic-scale description of CtErbB2, using both NMR and SAXS. We show that although this region is highly disordered, it exhibits partially formed α helices, and a long-range contact between its N- and C-termini. Several proline-rich regions also transiently adopt PPII helices including potential SH3- and WW-binding short linear motifs. Together, our data suggest the existence of yet unknown conformational regulation processes of ErbB2 function.

MATERIALS AND METHODS

Protein expression and purification

CtErbB2 was expressed and purified as described in our previous study (30). The final construct, depicted in Fig.1 A, comprises residues 988-1255 of the full-length human ErbB2 receptor (numbered here from 1 to 268), plus four additional N-terminal residues (Gly-Ser-His-Met, numbered from -4 to 0).

NMR spectroscopy

The typical samples for NMR studies were composed of 200-400 μM protein, in 200-250 μL of MES buffer pH 5.6 (40 mM MES, 200 mM NaCl, 2 mM TCEP, 5% D2O). NMR spectra were recorded at 298 K. Three different spectrometers, each equipped with a TCI cryoprobe (^{1}H, ^{13}C, ^{15}N, ^{2}H) and z-axis pulsed field gradients, were used: a Bruker Avance III 950 MHz spectrometer (22.3 T), a Bruker Avance III 800 MHz spectrometer (18.8 T), and a Bruker Avance III 600 MHz spectrometer (14.1 T) (Bruker, Billerica, MA, USA). Data were processed using TOPSPIN 3.5 (Bruker) and analyzed with CCPNMR (38).

^{15}N relaxation ^{15}N relaxation parameters R_1, R_2 and ^{15}N-$\{^1H\}$ steady-state nuclear Overhauser effects (NOEs) of CtErbB2 were measured at 14.1, 18.8 and 22.3 T (600, 800 and 950 MHz proton frequency, respectively), using pseudo-3D HSQC-based pulse sequences comparable to those implemented in the Bruker pulse sequence library. 10 to 14 relaxation delays were used, ranging from 20 ms to 2 s for R_1 measurements, and from 4 to 450 ms for R_2 measurements. Relaxation rates were obtained by fitting the decay curves to a two-parameter single exponential decay function. The ^{15}N-$\{^1H\}$ nOes were measured using a TROSY-based pulse sequence using 120° pulses every 5 ms for 4 s for proton saturation. The ^{15}N-$\{^1H\}$ nOe values were calculated as the intensity ratio of peaks in the experiment with or without proton saturation, that were acquired in an interleaved

manner.

The ^{15}N relaxation parameters were then compared with different models derived from the "model-free" approach (39) and from the Hall and Helfand polymer model (40). Least-square fitting for each residue was done in MATLAB (The MathWorks, Natick, MA, USA) with the trust-region-reflective algorithm. Goodness of fit was assessed calculating reduced χ^2 values for each residue:

$$\chi^2 = \frac{1}{n-p}\sum_f \left[\left(\frac{R_{1,f} - R_{1,f}^{calc}}{dR_{1,f}} \right)^2 + \left(\frac{R_{2,f} - R_{2,f}^{calc}}{dR_{2,f}} \right)^2 + \left(\frac{nOe_f - nOe_f^{calc}}{dnOe_f} \right)^2 \right]$$

where dR_1, dR_2 and $dnOe$ are the errors on the experimental relaxation parameters, n is the number of experimental parameters (here 9, 3 parameters measured at 3 magnetic fields), p the number of fitted parameters in the model, and f the magnetic field. To estimate errors on those fitted parameters, the standard deviation of 1000 Monte-Carlo runs were used. The fit was considered statistically satisfactory if the error function χ^2_{exp} obtained from experimental data (equation above) lay within the 95% confidence limit $\chi^2_{0.95\%}$ obtained from the 1000 Monte Carlo simulations, i.e. $\chi^2_{exp} < \chi^2_{0.95\%}$.

The model that fitted best the experimental data was the Hall and Helfand model modified by DeJean de la Batie, Lauprêtre and Monnerie (41, 42) (DLM model). The Hall and Helfand polymer model assumes two motional parameters, τ_1 and τ_2, associated with crankshaft-like transitions that consist of segmental motions by rotation around two collinear bonds, and that can be classified according to the relative position in the initial and transition states of the tails attached to the segment undergoing the transition. These transitions can be characterized by the number of conformational jump (s) and the amplitude of the motion of the tails it generates. Transition probability is then determined by the balance between the energy barriers for intermolecular bond rotation and the hydrodynamic forces hindering the motion of segments through the medium. A conformational transition induces distortions of the tails that relax by translations and/or rotations. It was shown from theoretical calculation and Brownian dynamics simulations that these distortions are mostly damped over a limited number of neighboring bonds in polymers (43, 44). The damping process consists of either non-propagative specific motions or distortions of the chain with respect to its most stable conformations. Moreover, it was found that the extent of distortion of the neighboring bonds is often so great as to lead to two simultaneous transitions. This latter process can be considered as a diffusion of bond orientation along the chain. In the case of proteins, correlated transitions correspond to conformational jumps $(\phi_{i-1}, \psi_i) \leftrightarrow (\phi_i, \psi_{i+1})$ ((45, 46). The energy involved in such transitions is controlled by steric interactions (Ramachandran plot), hydrogen bonds and electrostatic interactions. The DLM model additionally takes into account independant libration of the H-N bond, with the introduction of an order parameter A and an associated correlation time τ_f. In our case, the simple Hall and Helfand model did not yield good fit of the data. The addition of the order parameter A allowed a statistically relevant fit, which was not improved by the addition of the third correlation time τ_f. The spectral density function was finally :

$$J(\omega) = Re\left(\frac{A}{(\alpha + i\beta)^{1/2}} \right)$$

where

$$\alpha = \frac{1}{\tau_2^2} + \frac{2}{\tau_1 \tau_2} - \omega^2$$

and

$$\beta = -2\omega\left(\frac{1}{\tau_1} + \frac{1}{\tau_2} \right)$$

τ_1 is the characteristic time for the correlated bond orientation jumps along the chain while τ_2 describes the damping of the orientation propagation along the chain on each side of the conformational jumps.

Residual Dipolar Couplings measurements The sample was aligned in 6% acrylamide neutral gels. The protein sample was used to rehydrate the gel after dialysis in water and drying at room temperature. The gel was stretched into an open-ended 5mm NMR tube with the help of the apparatus described by Chou *et al.* (47). H_N-N ($^1D_{HN}$) dipolar couplings were measured using a BEST-HSQC-IPAP, at a proton frequency of 800 MHz, at 298 K. The whole procedure was repeated with a different protein sample and a different gel to estimate measurement errors. The respective deuterium quadripolar splittings were 4.98 and 5.66 Hz.

Experimental values were compared with those calculated by Flexible Meccano (48) on two ensembles of 10 000 structures generated without constrains (random coil) or with the following secondary structure propensities: an α helix at 20% between residues 15-20, PPII helices (defined with dihedral angles (-78°, 149°)) at 20% (residues 76-93, 178-191, 216-226 and 238-261), 30% (residues 111-122, 141-151 and 168-175) and 40% (residues 156-165).

Scalar couplings measurements The homonuclear $^3J_{HNHA}$ coupling constant for each residue was measured with the 3D HNHA experiment of Vuister and Bax (49) implemented by Bruker.

Paramagnetic Relaxation Enhancement Four mutants of CtErbB2 were used for four different experiments: C146S, C45S, (C45S,C146S,S227C) and (C45S,C146S,S248C) (for respective paramagnetic tag positions of C45, C146, C227 and C248). These ^{15}N-^{13}C labeled mutants were reduced with TCEP and incubated at 15°C overnight with the paramagnetic probe (1:10 protein:probe ratio) in a 20 mM sodium borate buffer at pH 8.0. The buffer was then exchanged for the MES buffer pH 5.6 (ZebaSpin Desalting columns, ThermoScientific), removing the excess of free paramagnetic probe. The ^{15}N-^1H HSQC spectrum of each paramagnetic form was recorded at 950 MHz. The probe coupling was total, as assessed by MALDI-TOF mass spectrometry (data not shown). To obtain a reference spectrum of the non-paramagnetic proteins, the samples were reduced with ascorbic acid and a second HSQC spectrum was acquired (with peak intensities I_0 in Fig.2 C). The assignment of each reference spectrum was performed by recording BEST-type HNCA, HNCACB and HNCOCACB (50) and HADAMAC (51) experiments. The probes that were used in this study are 3-(2-Iodoacetamido)-PROXYL (IAP) and S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL). The spectra of the probe-attached protein were highly similar to that of the apo-protein for both IAP and MTSL, indicating no major perturbation of the conformational ensemble. Only the results with MTSL, giving similar but stronger effects compared to IAP, are presented here (Fig.2 C).

Small angle X-ray scattering

The samples for SAXS were similar in composition to those used for NMR. The measurements were done on a Nanostar Instrument (Bruker) with a Microstar rotating anode (λ = 1.54 Å) at 285 K . Several curves were recorded at concentrations from 50 to 200 μM, and were identical, indicating the absence of oligomerization. A Sharp-Bloomfield (52, 53) equation was used to fit the data and obtain polymer parameters for the polypeptide chain (54, 55):

$$\frac{I(q)}{I(0)} = I_{SB}(q)e^{-q^2R_c^2/2}$$

where

- $I_{SB}(q) = g_D(x) + \frac{b}{L}\left(\frac{4}{15} + \frac{7}{15x} - (\frac{11}{15} + \frac{7}{15x})e^{-x}\right)$
- $g_D(x) = 2\frac{e^{-x}+x-1}{x^2}$ is the Debye function
- $x = q^2Lb/6$
- L is the contour length of the chain
- b is the length of the statistical element (twice the persistence length)
- R_c is the radius of gyration of cross section

The theoretical value of the contour length for an protein is calculated as $L = Nl_0f$ where N is the number of residues, l_0 is the distance between two sequential C_α, and f is a geometrical factor taking into account dihedral angles. For an unstructured protein we take l_0 = 3.78 Å and f = 0.95 as in (55).

Circular dichroism

The measurements were performed on a Jasco J-810 spectropolarimeter (Jasco Inc, Hachioji-shi, Tokyo, Japan) equipped with a sample cell temperature control unit (PFD 423S/L Peltier), in a 0.01 mm path length quartz cell. CtErbB2 (200 μM) was examined in the same MES buffer as used for NMR. The wavelength range was 190-260 nm, with a wave step of 0.1 nm. The spectra consisted of an average of ten scans acquired at a speed of 50 nm/min. Spectra were recorded at 298 K. The buffer contribution was subtracted and the signal normalized with protein concentration and number of residues to convert the results to molar ellipticity per residue units (MER).

RESULTS

CtErbB2 has no stable local or tertiary structure

The construct of CtErbB2 that we used in this study is presented in Fig.1 A. Circular dichroism of CtErbB2 was measured at 298 K and shows the characteristic signature of an IDP, with only one strong negative band at 200 nm (Fig.1 B). Small angle X-ray scattering (SAXS) also supports this observation. The scattering curve of CtErbB2 is given in Fig.1 C. The Kratky plot and the distance distribution are characteristic of a disordered protein, giving a radius of gyration of 49.2 Å from the P(r) distribution. Flory theory gives the radius of gyration R_g of a polymer with N monomers, $R_g = R_0 N^{\nu}$, where R_0 and ν depend on the behavior of the polymer in solution, and especially its solvation. Its applicability to denatured or disordered proteins has already been investigated (55). Bernadó and Blackledge (56) found that parameters $R_0 = 2.54$ Å and $\nu = 0.522$ recapitulated well the behavior of such chains. For CtErbB2, this equation gives $R_g = 47.4$ Å, consistent with our experimental value. It is to be compared with an estimate for a globular protein of the same molecular weight, that would give $R_g = 18.7$ Å, with $R_0 = \sqrt{3/5} * 4.75$ and $\nu = 0.29$ (57).

At atomic resolution, the poor dispersion of CtErbB2 proton resonances in the ^{1}H-^{15}N HSQC spectrum is typical of an intrinsically disordered protein (Fig.1 D). The ^{1}H, ^{13}C and ^{15}N assignment of CtErbB2 (backbone and side chains) was previously achieved (30) and the chemical shifts were deposited into the BMRB (entry 26740). $\delta(^{13}C_{\alpha})$, $\delta(^{13}C_{\beta})$ and $\delta(^{1}H_{\alpha})$ were used as input for the SSP algorithm (58), and the results are presented in Fig.2 A. The SSP (Secondary Structure Propensity) score is a quantification of secondary structure formation. SSP scores of 0 indicate a fully random coil conformation, while scores of -1 or +1 indicate fully stable secondary structures, either extended (-1) or compact (+1). The SSP scores mainly between -0.2 and 0.2 indicate the lack of fully stable secondary structures in CtErbB2. Moreover, backbone N-H residual dipolar couplings (RDCs) measured in stretched acrylamide gels ($^{1}D_{NH}$, Fig.2 A) are negative almost all along CtErbB2 sequence, as usually observed for denatured or disordered proteins (59, 60). Overall, all the data show that CtErbB2 is an intrinsically disordered region (IDR).

CtErbB2 displays local propensities for a residual α helix and numerous PPII helices

Many IDPs interact with globular partners through molecular recognition features (MoRFs), which are small sequences exhibiting transient secondary structures in the free state that are stabilized upon interaction. Although these MoRFs can be of all kinds, most of them are α helices (61). Given the high content of prolines in CtErbB2, polyproline helices are also likely to be locally populated. We sought to investigate whether CtErbB2 exhibited such transient secondary structures along its sequence, taking advantage of the residue-specific structural information given by NMR parameters, namely SSP scores, H_N-H_{α} J couplings ($^{3}J_{HNHA}$) and $^{1}D_{NH}$ RDC (Fig.2 A). The deviations of these parameters from expected values for residues in random coil conformation provide information on their propensity to adopt a specific secondary structure.

We firstly investigated the conformation of CtErbB2 numerous prolines. Polyproline helices of type II (PPII) are a common structural motif among disordered peptides and proteins, especially those rich in proline residues (62). These left-handed helices are characterized by an exclusively *trans*-proline conformation, as opposed to the prolines in random coil fragments, for which *cis* and *trans* conformations coexist, with a predominant trans conformation. We previously observed that for all but two prolines in CtErbB2 (indicated by grey bars in Fig.2 A), only *trans* conformation could be identified at our working concentrations (solid blue bars in Fig.2 A), which rules out the possibility of *cis* prolines formed at more than about 10%. To determine whether the absence of peaks for *cis* prolines was due to lack of sensitivity, HSQC spectra at higher concentrations were recorded, which showed many low-intensity peaks, some of which could be assigned to residues following cis prolines (shown in supplementary information, Fig. S1). However, due to still insufficient signal-to-noise ratio, no sequential assignment of these peaks could be achieved.

Typically, positive SSP scores indicate compact structures (3_{10}, α or PPI helices) which have less than 2 Å of translation per residue, while negative scores indicate extended structures (β strands or PPII helices) with a translation per residue typically higher than 3 Å.

In CtErbB2, such negative SSP scores are well correlated with strongly negative $^{1}D_{NH}$ values, and are predominantly observed in proline-rich stretches (indicated by vertical bars in Fig.2 A). This suggests the presence of PPII helices, rather than β strands. H_N-H_{α} J-couplings ($^{3}J_{HNHA}$) make it possible to differentiate these two structures. Indeed, residues involved in β strands have higher $^{3}J_{HNHA}$ (more than 9 Hz) than those involved in PPII helices, which exhibit $^{3}J_{HNHA}$ around 5-6 Hz, close to or below the random coil value (around 6-7 Hz). In Fig.2 A are presented $^{3}J_{HNHA}$ values, compared to random coil $^{3}J_{HNHA}$ predicted for this sequence (https://spin.niddk.nih.gov/bax/nmrserver/rc_3Jhnha/ (63)). Experimental $^{3}J_{HNHA}$ being almost always equal or below the predicted random coil value, no propensity for β strands can be detected in CtErbB2, in

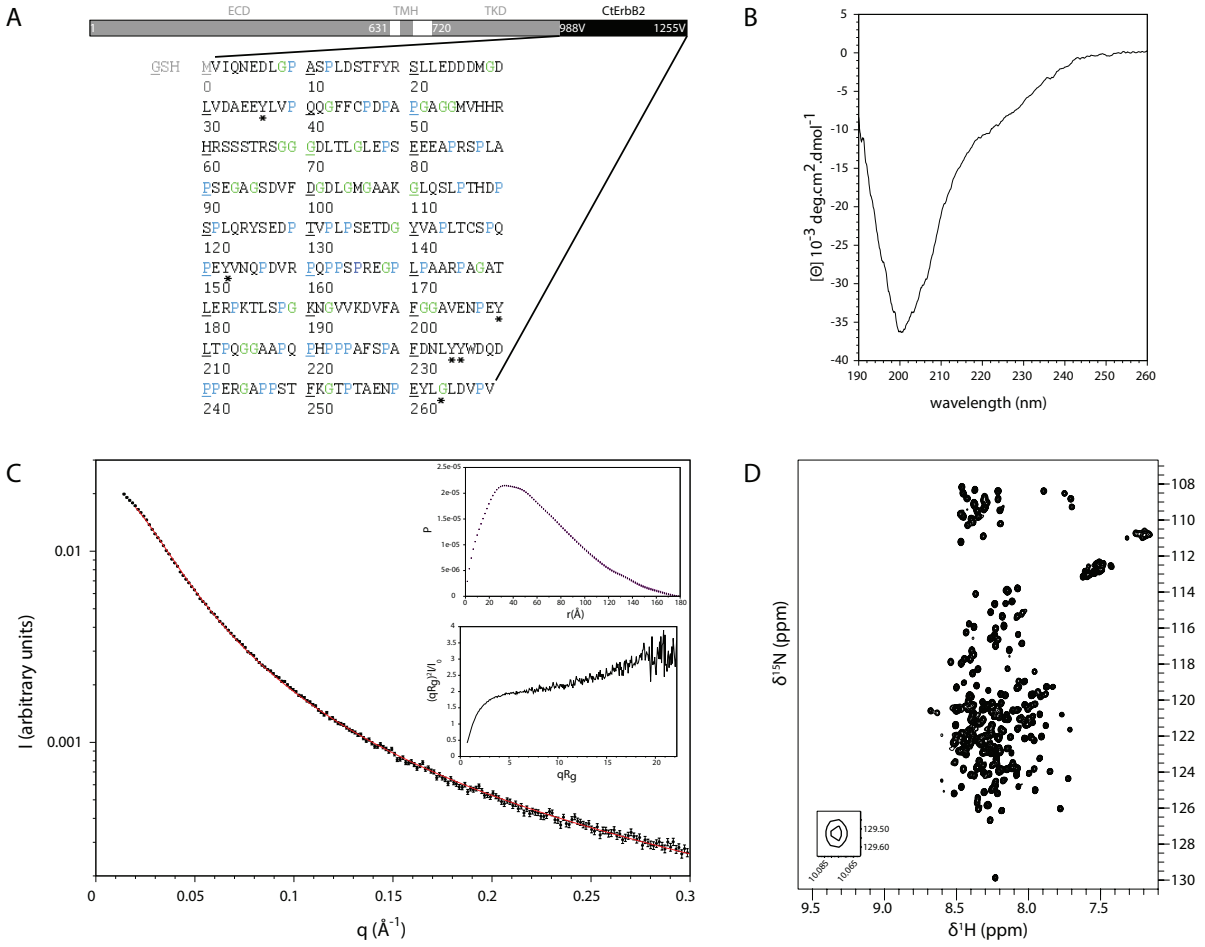


Figure 1: ***(A)*** Schematic representation of the human ErbB2 receptor domains and sequence of the CtErbB2 construct. ECD = extracellular domain; TMH = transmembrane helix; TKD = tyrosine kinase domain; CtErbB2 = C-terminal tail. Valine 1 in our CtErbB2 construct corresponds to valine 988 in the full length receptor. Asterisks indicate tyrosines that are phosphorylated by ErbB kinases in dimers of full receptors (autophosphorylation). ***(B)*** CD spectrum (mollar ellipticity per residue) of CtErbB2. ***(C)*** *Main* Experimental SAXS intensity (black dots) and fit with Sharp Bloomfield equation (red line). *Insets* Top: Distance distribution. Bottom: Kratky plot. ***(D)*** ^{1}H-^{15}N HSQC spectrum of CtErbB2 at 950 MHz.

extended regions or elsewhere. All our data indicate that the extended, proline-rich regions identified are PPII helices.

Positive SSP scores point to several zones with propensities to form compact helices, that are all N-capped by a serine or aspartate (^{14}DSTFYRSLL^{22}, ^{71}DLTLGL^{75}, ^{100}DGDLGMG^{105}, ^{135}SETDGY^{140}, ^{231}DNLYY^{235}). The first one is longer and more populated (around 20% according to SSP). It corresponds to higher $^{1}D_{NH}$ values, consistent with N-H bonds of α-helices partially aligned parallel to the gel stretching direction and the magnetic field. These N-capped helices are inserted between extended regions, giving an overall alternation of the two kinds of structures. The same kind of profile was already observed for the measles virus nucleoprotein N_{TAIL} region (64), in which additionally to the most populated (N-capped) helix, small stretches N-capped by aspartates have positive SSP scores and alternate with more extended stretches.

Globally, all NMR data indicate the presence of transiently populated secondary structures along the sequence of CtErbB2, with a N-terminal α helix structured at about 20% and PPII regions with propensities between 20 to 40%. The $^{1}D_{NH}$ RDCs back-calculated from 10000 conformers generated in the software Flexible Meccano (48) with no secondary structure propensities significantly deviate from the experimental values, whereas thoses obtained using these propensities as input constraints correctly account for the experimental values (orange and red lines in Fig.2 A, respectively).

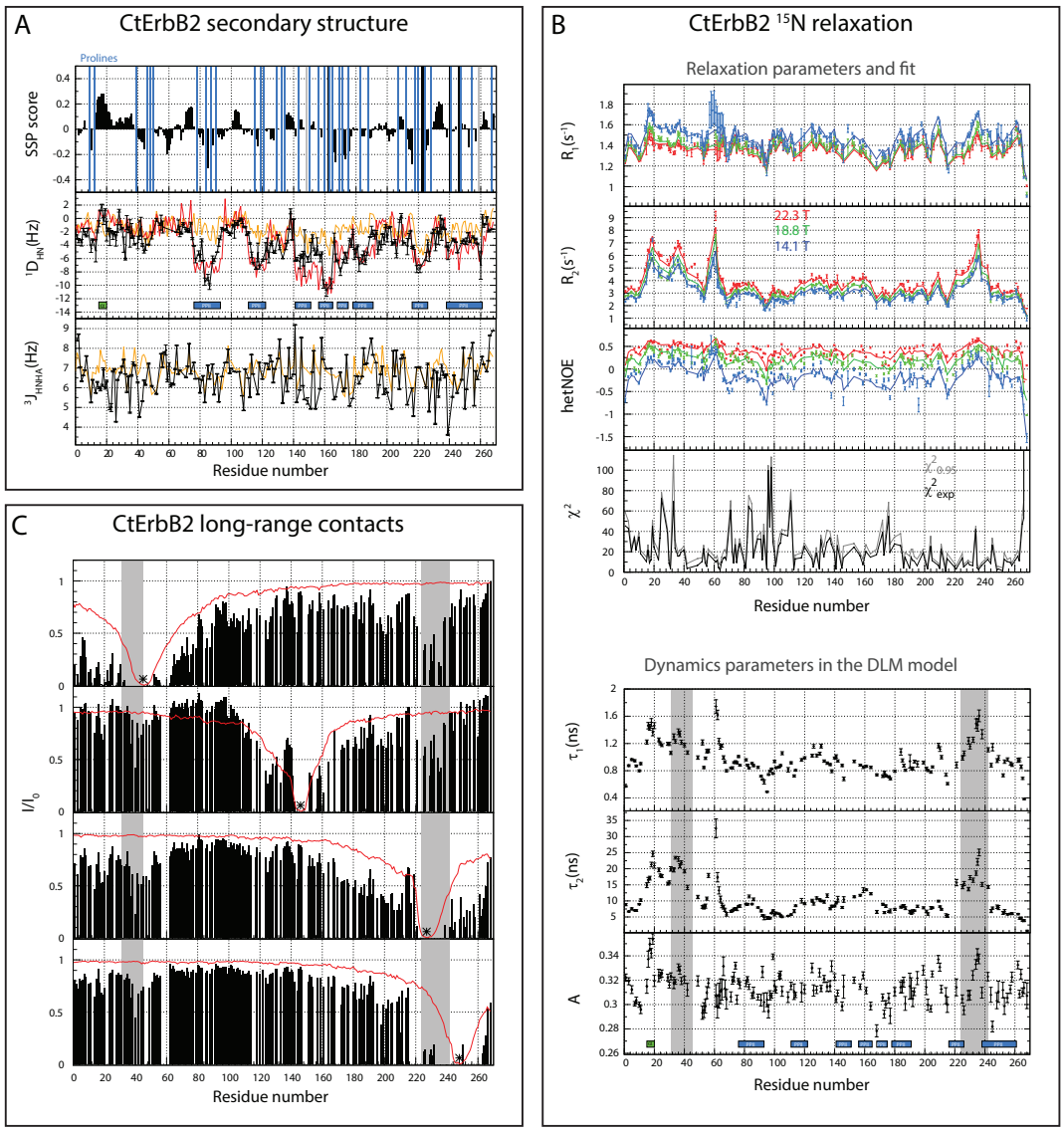


Figure 2: CtErbB2 secondary and tertiary structure identification. Green box= α helix; blue box = PPII helix. Grey area: zones of tertiary contact. (***A***) *Top*: SSP scores (58). Vertical bars represent the position of prolines. Blue bars: *trans* prolines (very few or no *cis* conformation detected); Grey bars: prolines with significant proportion of *cis* conformation detected; Black bars: prolines of unknown conformation. *Middle*: N-H residual dipolar couplings compared with Flexible Meccano (48) calculations. Black: experimental; orange: calculated from a random coil Flexible Meccano ensemble (10 000 structures); red: calculated from a random coil Flexible Meccano ensemble with the secondary structure propensities presented below. *Bottom*: $^3J_{HNHA}$. Black : experimental; orange: calculated for a random coil (63). (***B***) Upper pannel: Experimental (points) and recalculated (lines) ^{15}N relaxation parameters of CtErbB2 at three magnetic fields using the DLM model described in the Materials and Methods section. χ^2 values were calculated for experimental versus recalculated data (χ^2_{exp}) and for 95% lower-χ^2 Monte-Carlo simulations versus recalculated data ($\chi^2_{0.95}$) ; lower pannel: Dynamics parameters of CtErbB2 determined from fitting $\{^{15}N\}R_1$, $\{^{15}N\}R_2$ and heteronuclear nOes data at three magnetic fields to the DLM polymer model (see Materials and Methods). Missing data correspond to prolines and weak, unassigned or overlapping peaks. (***C***) Paramagnetic relaxation enhancement measurements for mutants C146S, C45S, (C45S,C146S,S227C) and (C45S,C146S,S248C) of CtErbB2. The paramagnetic probe, indicated by an asterisk, was attached to cysteine 45, 146, 227 and 248, respectively. The red line is the PRE values predicted for the ensemble generated by Flexible Meccano taking into account secondary structures.

ErbB2 exhibits heterogeneous dynamic behavior along its backbone and some level of compaction

NMR, and in particular ^{15}N spin relaxation parameters, makes it possible to highlight restriction of dynamics compared to a random coil. The presence of transient secondary structure is expected to locally alter the reorientation dynamics of NH bounds

in the 100 ps to nanoseconds timescale that is responsible for ^{15}N spin relaxation. Moreover, transient long-range interactions are likely to occur, as part of fly-casting mechanisms (33) and regulation of partner accessibility. The transient modification of the electronic environment due to these long range contacts can also induce chemical exchange and an increase of ^{15}N transverse relaxation rates ^{15}N R_2 if they occur in the millisecond time scale.
^{15}N R_2 and R_1 and $\{^1H\}$-^{15}N nOe values, measured at three different magnetic fields (14.1, 18.8 and 22.3T corresponding to 600, 800 and 950 MHz proton frequencies) for CtErbB2, are depicted in the upper pannel of Fig.2 B. The bell-shaped profile that is expected for a completely disordered chain is very distorted, indicating a heterogeneous dynamic behavior, with short fragments exhibiting different type of flexibility (65). To gain more insight into CtErbB2 dynamics behavior, the data were first analyzed using the classically used "model-free" approaches, which didn't yield statistically relevant fits. Data could be properly fitted with the polymer model developed by DeJean de la Batie, Lauprêtre and Monnerie (41, 42), abbreviated the DLM model. This model, initially designed to describe the dynamics of polymers in solution, appears to be well suited for a long, globally disordered protein chain such as CtErbB2. Our analysis allowed extracting of three dynamic parameters (lower panel of Fig.2 B): an order parameter revealing the amplitude of libration of individual N-H vectors (in a timescale of around 10 ps), τ_1, characteristic of the diffusion of bond orientation along the chain and τ_2, corresponding to the damping of the orientation propagation, i.e. the orientation loss processes. The timescales covered by this approach (tens of ps for reorientation of vectors, low ns local flexibility, and a few to tens of ns when local structure formation is observed) cover the main ranges recently identified as typical for IDPs using the more complex *IMPACT* approach (66).
Along the chain of CtErbB2, low and homogeneous order parameter values A (0.31 ± 0.02) indicate a high proportion of fast, high amplitude, librations motions of NH bonds in the tens of picosecond range, that depends neither on the nature of the residue nor on its implication in a secondary structure. On a slower timescale, the damping of the transitions, indicated by the second correlation time τ_2, exhibit much larger variations. It ranges roughly between 5 and 20 ns (10.4±5 ns), and overall follows the trend of τ_1 (local concerted transitions of bounds orientation), which lies between 0.38 and 1.75 ns (1.00±0.23 ns). Four different types of behaviors can be qualitatively distinguished :

- (i) In the N and C-terminal ends and the glycine rich segment (65-110), τ_1 is around 1 ns or below, typical of a fully random coil, with minimal values of damping correlation times τ_2 around 5 ns. In the middle of the chain, the lowest values of τ_2 are interestingly found near glycines repetitions ($G_{68}G_{69}G_{70}$, $G_{201}G_{202}$, $G_{214}G_{215}$, or glycine rich segment ([93-110] with 6 glycines).

- (ii) The histidine rich region (56-61) exhibits much more restricted motions with highest τ_1 and τ_2 (almost 2 and 35 ns respectively). Hydrogen bonds can hinder the flexibility of this highly polar stretch, which also contains several arginines and serine.

- (iii) The two segments (15-49) and (225-244) also exhibit high values of both τ_1 and τ_2, indicating a combination of restricted motions at different timescales and fast motions below 1ns. The transient presence of the N-terminal helix can explain this restricted flexibility around the (15-20) segment, with also sightly restricted libration (increased value of A). The other two regions (31-49 and 225-244) display only small helical propensities, and therefore their peculiar dynamic features can not be explained only with secondary structure considerations. Transient tertiary structure (long-range contacts) may be at stake here, that could impeed the damping processes of the bounds orientational transitions. To determine whether this rigidity in the ns timescale was associated with slower dynamics (μs to ms), CPMG experiments were recorded and showed no sign of exchange in this timescale except for rather homogeneous exchange along the sequence attributed to NH exchange with water (data not shown).

- (iv) In the rest of the protein, τ_1 and τ_2 values indicate motions in the low to high ns timescales slightly more restricted than in the glycine rich region. Typically, this encompasses PPII regions, that define dynamically independent stretches separated by more flexible, glycine-containing nodes (green in Fig.1 A). These stretches are often less than ten residues long, not very different from the length of the statistical element estimated for a completely disordered chain (65, 67, 68), explaining the dynamical behavior close to that of a random coil.

SAXS is a method complementary to NMR when it comes to assessing degree of secondary and tertiary structure of a disordered protein (37). To obtain detailed information from the scattering curve, we fitted the intensity to the Sharp-Bloomfield equation, using a polymer model, as previously done for a Repeat-in-ToXin domain (54). This model (red curve in Fig.1 C) yielded good reproduction of the data with three fitted parameters: the length of the statistical element b ~ 20 Å (which gives a persistence length of 10 Å), the contour length L ~ 770 Å and the radius of gyration of cross section (thickness of the chain) R_C = 2.7 Å. The 10 Å persistence lenth corresponds to a six amino-acids statistical chain, close to the value of seven estimated by Schwalbe et al.(68) for a chain with few, or short, secondary structure elements. However, the value of the contour length,

which reflects the length of the protein when extended without disrupting secondary structures, seems low compared to what is expected for a disordered chain of 272 residues (around 980 Å, see experimental procedures for details of the calculation). This is compatible with the presence of some residual local structure. The presence of many prolines could also contribute to this short contour length, since even low-populated *cis* prolines have an effect, via reduced C_α- C_α distances.
The effect of structure formation on the radius of gyration is more complex, since the latter can be modified by both secondary and tertiary structures. Comparison of experimental R_g with those of simulated conformational ensembles is a useful tool to assess the different contributions. Here, the random coil ensemble generated by Flexible Meccano has an average R_g similar to the experimental one, 50.2 Å. However, when secondary structure propensities are taken into account, and due mainly to the extended nature of PPII helices, this value increases to 57.0 Å, substantially higher than the measured one. This increase in R_g could be counterbalanced by long-range interactions compacting the protein.
All in all, dynamic studies by NMR and SAXS data show that transient long-range contacts exist in the conformational ensemble of CtErbB2.

PRE suggests a N-to-C terminal contact

To investigate the existence and location of tertiary contacts in the conformational ensemble of CtErbB2, especially any contact that would explain the specific dynamic behavior of CtErbB2 terminus regions, paramagnetic relaxation enhancement (PRE) was measured. We used four mutants, C146S, C45S, (C45S, C146S, S227C) and (C45S, C146S, S248C), each permetting coupling of MTSL on a single cystein (the two native cysteines and two mutated serines). This mutant collection enables detection of any contact involving the N-terminus, central region or C-terminus of CtErbB2. The I/I_0 values, were I is the intensity of each HSQC peak when the probe is paramagnetic, and I_0 when the probe is reduced (diamagnetic) are presented in Fig.2 C, together with the values predicted for the ensemble with secondary structures generated with Flexible Meccano. When the paramagnetic probe is attached to the N-terminus, a decrease in intensity is observed at the C-terminus and *vice-versa*, with a stronger effect when the probe is attached to residue 227 compared to 248. This contact is consistent with high R_2 observed in the terminal regions, as highlighted in grey in Fig.2 B.
Other contacts are observed between the central part and the extremities of CtErbB2 when the probe is attached at position 146, but they are not observed reciprocally, which could indicate hydrophobic contacts induced by the probe. In agreement with this interpretation is the fact that the region around residues 190-210 is the most hydrophobic in the sequence, and shows paramagnetic relaxation enhancement in all experiments. Globally, these experiments thus strongly suggest a long-range contact between the two segments 30-40 and 220-240 of CtErbB2.

Conservation of functionally and/or structurally important elements of CtErbB2 amongst mammals

We sought to link our observations on structure and dynamics of CtErbB2 to functional features by looking at the sequence conservation of CtErbB2 amongst mammals. Since conservation of some of the characteristics of CtErbB2 may not require strict sequence conservation but, for example, conservation of high proline density or conservation of interaction motifs, we analyzed sequence conservation node by node on a tree made with mammalian sequences selected with PSI-BLAST. This allowed depiction of more subtle changes, such as displacements of polyproline motifs, compared to analyzing single consensus sequences. The method, as well as the resulting alignments are given in the Supporting Material (supplementary experimental procedures and Fig. S2-S5). The 6 tyrosine autophosphorylation sites are strictly conserved amongst mammals, as expected given their role in signal transduction. Despite the high level of flexibility of CtErbB2, a feature that is often associated with low sequence conservation, the whole region is well conserved amongst mammals even outside these sites (one third of the human sequence is strictly conserved, 15% has a score higher than 0.5 as calculated by Clustal Omega). The same observation was made for EGFR (69). Other structural characteristics of CtErbB2 are then likely to bear functional consequences. Interestingly, the first 50 residues are particularly well conserved (31 are strictly conserved, 10 have scores > 0.5). This region contains the transient N-terminal helix, as well as the N-terminal region involved in long-range transient contacts. The C-terminal region involved in this long-range contact is the second best conserved region (the ^{228}PAFDNLYYW^{236} stretch is strictly conserved). Finally, in the PPII regions and PxxP motifs, even though sequence conservation is not strict, proline density is conserved overall, and some mutations of prolines around PxxP sequences compensate each other thus leading to the conservation of the motif (PxxP motifs 6 and 7).

DISCUSSION

CtErbB2 disorder is relevant in the context of the whole receptor

All performed experiments indicate that CtErbB2 is mainly devoid of persistent secondary or tertiary structure. CD gives no sign of the presence of stable secondary structure in the protein. Deviation from random coil chemical shifts gives locally no more than 20% of secondary structure, which is rather low even for transiently formed structures in IDPs. Many regions of the protein adopt an extended conformation, but are not drastically constrained in their dynamic behavior. This is consistent with the model of an anisotropic, polymer-like random coil model, where excluded volume effects favor extended conformations (59, 60, 68). This extended nature of disorder gives rise to negative N-H RDCs in stretched gels (59, 60).

Evidence suggest that the disordered behavior observed here is a feature of CtErbB2 in the context of the full-length protein. The other domain that is most likely to interact with CtErB2 is the kinase domain. Keppel *et al.* showed by hydrogen-deuterium exchange followed by mass spectrometry that CtErbB2 is still disordered even in the context of the whole cytoplasmic region (31). Moreover, the disordered nature of CtErbB2 is consistent with its function within the full receptor, which is based on accessibility of interaction sites: its tyrosine side-chains need to be accessible to the kinase domain in the context of activated dimers.

The N-terminal helix may contribute to regulation of the kinase activity of ErbB2

Despite the disordered nature of CtErbB2, residual helical secondary structure does exist. It is of interest since many IDPs undergo folding upon binding, mostly in α helices that are sometimes preformed (70–72). The longest and most populated helix in CtErbB2 is located between residues 14 and 22 (DSTFYRSLL), and contains one tyrosine, which has not been reported to be an interaction site for adaptor proteins *in vivo*. However, in the study by Keppel *et al.* (31), this region was shown to be more solvent-protected than the rest of the C-terminal tail. Furthermore, this helix is visible in the ErbB2 kinase crystal structures, in which it is packed against the kinase domain (PDB 3RCD (23) and 3PP0 (22)). The helix, preformed independently from the kinase, could therefore be stabilized by intramolecular interaction.

Very similar interactions are observed between the kinase and C-terminal tail of EGFR, for which it has been shown that the first dozens of residues of the C-terminal tail have an inhibitory role on the kinase, both from structural and enzymatic data (29, 69, 73) and in terms of transformation potency (74). The involved EGFR helix is called the AP-2 helix, for the FYRAL motif has been shown to interact with the clathrin adaptor protein complex 2 (75). Despite overall moderate sequence conservation between the four ErbB proteins (42.0% similarity maximum, between ErbB2 and EGFR), the region of the helix is well conserved between EGFR (PSPTDSNFYRALMDEED) and ErbB2 (ASPLDSTFYRSLLEDDD), as it is the case for the 50 first residues of the tail (with similarity of 72.5%, see Fig. S5). This could support the hypothesis of a similar kinase-inhibitory role for these residues of CtErbB2. A tyrosine phosphorylation was suggested to disrupt this inhibitory intramolecular interaction in EGFR (29).

CtErbB2 has long-range contacts involving functionally important regions

The presence of at least one long-range contact is suggested by the relatively small R_g derived from SAXS compared to an ensemble with the determined secondary structure propensities. PRE experiments also show evidence of a contact between the N- and C- termini, with decreased intensity in one region when the MTSL probe is attached to the other, and vice-versa. Relaxation data confirm that these regions have a particular dynamic behavior. They notably contain less prolines and more hydrophobic and aromatic residues than the rest of the sequence, and are predicted to partially fold in helices (Fig. 2 A). It is not unusual in disordered proteins to observe coincidental secondary and tertiary folding (76). Whether this contact has functional importance remains to be investigated, but its overlap with important sites for signaling is to be noticed. It could affect accessibility of several phosphorylation and interaction sites, such as tyrosine 36 (Y1023 in the full-length protein, the phosphorylation of which was shown to be inhibitory of ErbB2-mediated transformation potency) and tyrosines 234-235 (Y1221-1222 in the full-length protein, an interaction site for Shc) (24). This contact could also be modified by phosphorylation in the activated receptor. In EGFR, an increase in the hydrodynamic radius of the intracellular domain was observed upon phosphorylation (77), correlating with increased dynamics of the C-terminus (29) and disruption of its interaction with the kinase domain (28). Whether CtErbB2 behaves similarly remains to be investigated.

CtErbB2 contains potential SH3 and WW binding motifs

The PPII (for PolyProline II) conformation is widely spread in proteins, especially in disordered and denatured ones, and in regions containing prolines(62). It was estimated that about 5% of residues (not only prolines) adopt this left-handed, extended triangular helical conformation of 3 residues per turn (78). PPII is still regularly ignored in secondary structure analyses, suggesting an even higher proportion. Given the lack of intra-helix hydrogen bonds and the high solvent accessibility of PPII helices, they are often stabilized by protein-protein interaction. They are mainly involved in interactions with the small SH3 and WW domains. Such interactions can be found in signaling pathways, typically those mediated by receptor tyrosine kinases (79), and WW domains are sometimes considered as the phosphorylation-dependent equivalent of SH3 domains (80). The consensus PPII interaction motif with SH3 domains is PxxP, with +xxPxxP and PxxPx+ (+ being either K or R) defining the most common class I and class II motifs (81). The main WW interaction motifs are PPxY (group I, with an unphosphorylated tyrosine), PPLP (group II), PPR or other proline-rich and K/R containing sequences (group III, potentially in competition for binding with SH3), and phospho-SP or phospho-TP (group IV)(80).

CtErbB2 is strikingly proline-rich (16% of residues are prolines). We could thus expect to find PPII helices in this tail. Negative SSP scores correlated with low $^{1}D_{NH}$ and $^{3}J_{HNHA}$ in *trans*-proline rich regions shows that the PPII conformation is indeed sampled at multiple positions along the CtErbB2 sequence. These helices are not especially rigid in the ps-ns timescales, and are only populated to a maximum of 40%, but could adopt a more stable conformation upon interaction. Three PPII stretches contain PxxP motifs : ^{81}EEA**P**RS**P**LA**P**SEGA^{94} (2 sequential motifs sharing one proline), ^{156}PDVR**P**Q**PP**S**P**REGP^{169} (two nested motifs, RxxPxxP and PxxPxxR), and ^{216}AAPQ**P**HP**P**PAFS^{227}. The first stretch is highly conserved in mammals (stretch 2-3 in Fig. S2). For the second stretch (number 5-6 in Fig. S2) , the 2 nested motifs are not fully conserved, but another PXXP or PXXPXXP motif, with prolines at position 169, 172 and 175, is frequently present in the same PPII rich regions in non human mammalian sequences. Two other PxxP motifs are found: ^{5}EDLG**P**AS**P**LDST^{16}, that is highly conserved among mammals (number 1 in Fig. S2), and ^{125}YSED**P**TV**P**LPSE^{136}. This last motif is in most non-human sequences part of a PXXPXXP motif with the last proline at position 137 (stretch 4 in Fig. S2). The isolated peptide ^{159}RPQPPSPRE^{167} studied by Bornet *et al.* (82) was already shown to adopt a PPII structure, and to be able to interact with FynSH3. One study also suggested that Grb2 SH3 domains bind to ErbB2 (83). Indeed, Grb2 SH2 domain (which is known to anchor Grb2 to CtErbB2 via tyrosine 152) does not recapitulate the binding of full length Grb2 to ErbB2.

There is no consensus PPxY, PPLP or PPR motif in CtErbB2, but the proline-rich regions with R or K residues constitute potential binding sites for group III WW domains. There are also several potential binding sites for group IV WW domains: seven SP motifs (residues 11-12, 86-87, 120-121, 147-148, 164-165, 187-188 and 227-228), and two TP motifs (residues 211-212 ans 253-254). Although a direct interaction between a WW domain and ErbB2 has never been observed, WW-containing proteins such as Pin1 and WWP1 have been shown to have a role in regulation of ErbB2-dependent pathways(84, 85).

CtErbB2 is known to ensure signal transduction thanks to interactions between its phosphorylated tyrosines and SH2/PTB/MEMO domains (4, 25, 26). PPII structures and proline-containing motifs expand the number of possible molecular recognition features of CtErbB2, with potential recognition by SH3 or WW domains.

CtErbB2 contains two prolines with significant *cis* population

Notable exceptions among prolines are P148 and P259, for which HNCOCACB peaks showing significant *cis* population are detected. However, it is probable that peaks corresponding to a *trans* population are present but superimposed with others, and therefore not assigned. Being located in positions 148 (^{146}CS**P**QPEYVN^{154}) and 259 (^{257}EN**P**EYLG^{263})), the *cis* prolines are close to tyrosines 152 and 261 that were shown to be phosphorylated and important for function (24, 27, 86). Additionally, the phospho-SP sequence of proline 148 could be a binding site for Pin1 prolyl-isomerase, modifying the *cis-trans* ratio. There could therefore be an interplay between proline isomerism and phosphorylation. No sequence or structure determinant can explain the presence of these populated *cis* prolines.

CONCLUSION

The characteristics of CtErbB2 described here are consistent with several reported functions of the disordered tail in the receptor: disorder ensures high accessibility and adaptability for post-translational modifications and multiple interaction with partners; a long-range contact could modulate accessibility; a preformed α helix suggests a regulation mechanism of the kinase activity by the tail; and PPII helices enable possible interactions with SH3 or WW containing proteins. These structural and functional features are summarized in Fig. 3, together with parts of the conformational ensemble generated by Flexible Meccano taking into account secondary structures. CtErbB2 can now be studied in other functionally relevant conditions, investigating

in particular how these characteristics are modified by tyrosine phosphorylation, and their impact on the protein interaction network and the regulation of signal transduction.

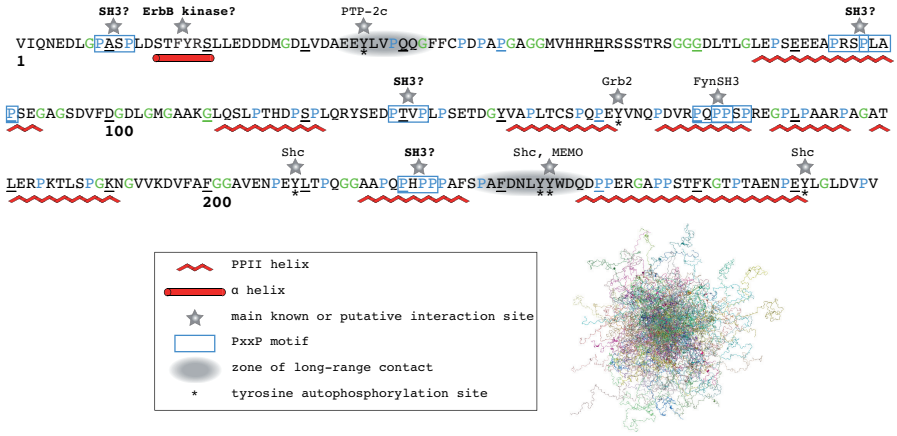


Figure 3: Model of CtErbB2 structural and functional features. *Along the sequence:* The transient secondary structures revealed by SSP scores, residual dipolar couplings and scalar couplings are indicated, as well as the main long-range contact observed in the PRE experiments. In addition to the main known interaction sites and corresponding partners (Shc and Grb2 (24, 25), PTP-2c (25) and MEMO(26) for interaction with phosphotyrosines, FynSH3 (82) for interaction with a RxPxxP motif) the PxxP motifs located in PPII helices are indicated as the best defined putative interaction sites of the proline-rich regions. The putative interaction of the N-terminal helix with the kinase domain is also indicated. *Samples of a conformational ensemble:* Superimposition of 100 structures randomly chosen from the ensemble of 10 000 structures generated by Flexible Meccano with transient secondary structures as input. The structures were aligned on residues -3 to 20.

AUTHOR CONTRIBUTIONS

Protein expression and purification was optimized and carried out by NA, YHW and LP. NMR experiments were optimized and recorded by LP, YHW, CvH, CD, EL and NM, and analyzed by YHW, LP, CvH and CD. CD experiments were recorded and analyzed by NA and LP. PRE design and experiments were performed by NA, YHW and LP. SAXS experiments were recorded and analyzed by DD. The amino-acid conservation study was performed by FB. CvH, AB and FG conceived the project, CvH, NA and FG supervised research, and CvH managed funding. LP and FG wrote the first draft of the manuscript. All authors discussed results and contributed to the final manuscript.

ACKNOWLEDGMENTS

We thank Jean-Pierre Le Caer and Vincent Guérineau for mass spectrometry verification of protein purification and probe coupling for PRE measurements, and Aurélien Thureau for help with complementary SAXS measurements. We also thank Arthur Besle and Anaïs Vogel for help in protein preparation and Annie Moretto for buffer, reagents and media preparation. This work was supported by the French National Research Agency (ANR, research grant ANR-13-BSV8-0016 to CvH, FG and AB and financing YW) and the French infrastructure for integrated structural biology (FRISBI). Louise Pinet had a Ph.D grant from the Université Paris-Sud. Financial support from the IR-RMN-THC Fr3050 CNRS for conducting the research is gratefully acknowledged.

REFERENCES

1. Lemmon, M. A., and J. Schlessinger, 2010. Cell Signaling by Receptor Tyrosine Kinases. *Cell* 141:1117–1134.

2. Roskoski, R., 2004. The ErbB/HER receptor protein-tyrosine kinases and cancer. *Biochemical and Biophysical Research Communications* 319:1–11.

3. Lax, I., A. K. Mitra, C. Ravera, D. R. Hurwitz, M. Rubinstein, A. Ullrich, R. M. Stroud, and J. Schlessinger, 1991. Epidermal growth factor (EGF) induces oligomerization of soluble, extracellular, ligand-binding domain of EGF receptor: A low resolution projection structure of the ligand-binding domain. *Journal of Biological Chemistry* 266:13828–13833.

4. Yarden, Y., and M. X. Sliwkowski, 2001. Untangling the ErbB signalling network. *Nature Reviews Molecular Cell Biology* 2:127–137.

5. Garrett, T. P., N. M. McKern, M. Lou, T. C. Elleman, T. E. Adams, G. O. Lovrecz, M. Kofler, R. N. Jorissen, E. C. Nice, A. W. Burgess, and C. W. Ward, 2003. The Crystal Structure of a Truncated ErbB2 Ectodomain Reveals an Active Conformation, Poised to Interact with Other ErbB Receptors. *Molecular Cell* 11:495–505.

6. Slamon, D. J., G. M. Clark, S. G. Wong, W. J. Levin, A. Ullrich, and W. L. McGuire, 1987. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science (New York, N.Y.)* 235:177–82.

7. Badache, A., and A. Gonçalves, 2006. The ErbB2 signaling network as a target for breast cancer therapy. *Journal of Mammary Gland Biology and Neoplasia* 11:13–25.

8. Parakh, S., H. K. Gan, A. C. Parslow, I. J. Burvenich, A. W. Burgess, and A. M. Scott, 2017. Evolution of anti-HER2 therapies for cancer treatment. *Cancer Treatment Reviews* 59:1–21.

9. Copeland, A. C., and C. K. Anders, 2018. Dual HER2-Targeting in the Adjuvant Setting: Where We Have Been and Where We Are Going. *Oncology (Williston Park, N.Y.)* 32:483–487.

10. Hu, S., Y. Sun, Y. Meng, X. Wang, W. Yang, W. Fu, H. Guo, W. Qian, S. Hou, B. Li, Z. Rao, Z. Lou, and Y. Guo, 2015. Molecular architecture of the ErbB2 extracellular domain homodimer. *Oncotarget* 6:1695–1706.

11. Franklin, M. C., K. D. Carey, F. F. Vajdos, D. J. Leahy, A. M. De Vos, and M. X. Sliwkowski, 2004. Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. *Cancer Cell* 5:317–328.

12. Zhou, H., Z. Zha, Y. Liu, H. Zhang, J. Zhu, S. Hu, G. Shen, L. Cheng, L. Niu, M. I. Greene, M. Teng, and J. Liu, 2011. Structural insights into the down-regulation of overexpressed p185her2/neuprotein of transformed cells by the antibody chA21. *Journal of Biological Chemistry* 286:31676–31683.

13. Fisher, R. D., M. Ultsch, A. Lingel, G. Schaefer, L. Shao, S. Birtalan, S. S. Sidhu, and C. Eigenbrot, 2010. Structure of the Complex between HER2 and an Antibody Paratope Formed by Side Chains from Tryptophan and Serine. *Journal of Molecular Biology* 402:217–229.

14. Lobner, E., A. S. Humm, K. Göritzer, G. Mlynek, M. G. Puchinger, C. Hasenhindl, F. Rüker, M. W. Traxlmayr, K. Djinović-Carugo, and C. Obinger, 2017. Fcab-HER2 Interaction: a Ménage à Trois. Lessons from X-Ray and Solution Studies. *Structure* 25:878–889.e5.

15. Cho, H.-S., K. Mason, K. X. Ramyar, A. M. Stanley, S. B. Gabelli, D. W. Denney, and D. J. Leahy, 2003. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* 421:756–760.

16. Bostrom, J., S.-F. Yu, D. Kan, B. A. Appleton, C. V. Lee, K. Billeci, W. Man, F. Peale, S. Ross, C. Wiesmann, and G. Fuh, 2009. Variants of the Antibody Herceptin That Interact with HER2 and VEGF at the Antigen Binding Site. *Science* 323:1610–1614.

17. Jost, C., J. Schilling, R. Tamaskovic, M. Schwill, A. Honegger, and A. Plückthun, 2013. Structural basis for eliciting a cytotoxic effect in HER2-overexpressing cancer cells via binding to the extracellular domain of HER2. *Structure (London, England : 1993)* 21:1979–91.

18. Eigenbrot, C., M. Ultsch, A. Dubnovitsky, L. Abrahmsen, and T. Hard, 2010. Structural basis for high-affinity HER2 receptor binding by an engineered protein. *Proceedings of the National Academy of Sciences* 107:15039–15044.

19. Bocharov, E. V., K. S. Mineev, P. E. Volynsky, Y. S. Ermolyuk, E. N. Tkach, A. G. Sobol, V. V. Chupin, M. P. Kirpichnikov, R. G. Efremov, and A. S. Arseniev, 2008. Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state. *Journal of Biological Chemistry* 283:6950–6956.

20. Mineev, K. S., E. V. Bocharov, Y. E. Pustovalova, O. V. Bocharova, V. V. Chupin, and A. S. Arseniev, 2010. Spatial Structure of the Transmembrane Domain Heterodimer of ErbB1 and ErbB2 Receptor Tyrosine Kinases. *Journal of Molecular Biology* 400:231–243.

21. Bragin, P. E., K. S. Mineev, O. V. Bocharova, P. E. Volynsky, E. V. Bocharov, and A. S. Arseniev, 2016. HER2 Transmembrane Domain Dimerization Coupled with Self-Association of Membrane-Embedded Cytoplasmic Juxtamembrane Regions. *Journal of Molecular Biology* 428:52–61.

22. Aertgeerts, K., R. Skene, J. Yano, B. C. Sang, H. Zou, G. Snell, A. Jennings, K. Iwamoto, N. Habuka, A. Hirokawa, T. Ishikawa, T. Tanaka, H. Miki, Y. Ohta, and S. Sogabe, 2011. Structural analysis of the mechanism of inhibition and allosteric activation of the kinase domain of HER2 protein. *Journal of Biological Chemistry* 286:18756–18765.

23. Ishikawa, T., M. Seto, H. Banno, Y. Kawakita, M. Oorui, T. Taniguchi, Y. Ohta, T. Tamura, A. Nakayama, H. Miki, H. Kamiguchi, T. Tanaka, N. Habuka, S. Sogabe, J. Yano, K. Aertgeerts, and K. Kamiyama, 2011. Design and synthesis of novel human epidermal growth factor receptor 2 (HER2)/epidermal growth factor receptor (EGFR) dual inhibitors bearing a pyrrolo[3,2-d]pyrimidine scaffold. *Journal of Medicinal Chemistry* 54:8030–8050.

24. Dankort, D. L., Z. Wang, V. Blackmore, M. F. Moran, and W. J. Muller, 1997. Distinct tyrosine autophosphorylation sites negatively and positively modulate neu-mediated transformation. *Molecular and cellular biology* 17:5410–5425.

25. Schulze, W. X., L. Deng, and M. Mann, 2005. Phosphotyrosine interactome of the ErbB-receptor kinase family. *Molecular Systems Biology* 1:E1–E13.

26. Feracci, M., C. Pimentel, O. Bornet, P. Roche, D. Salaun, A. Badache, and F. Guerlesquin, 2011. MEMO associated with an ErbB2 receptor phosphopeptide reveals a new phosphotyrosine motif. *FEBS letters* 585:2688–92.

27. Dankort, D., N. Jeyabalan, N. Jones, D. J. Dumont, and W. J. Muller, 2001. Multiple ErbB-2/Neu Phosphorylation Sites Mediate Transformation through Distinct Effector Proteins. *Journal of Biological Chemistry* 276:38921–38928.

28. Lee, N. Y., and J. G. Koland, 2005. Conformational changes accompany phosphorylation of the epidermal growth factor receptor C-terminal domain. *Protein science : a publication of the Protein Society* 14:2793–803.

29. Lee, N. Y., T. L. Hazlett, and J. G. Koland, 2006. Structure and dynamics of the epidermal growth factor receptor C-terminal phosphorylation domain. *Protein Science* 15:1142–1152.

30. Wang, Y., L. Pinet, N. Assrir, L. Elantak, F. Guerlesquin, A. Badache, E. Lescop, and C. van Heijenoort, 2018. 1H, 13C and 15N assignments of the C-terminal intrinsically disordered cytosolic fragment of the receptor tyrosine kinase ErbB2. *Biomolecular NMR Assignments* 12:23–26.

31. Keppel, T. R., K. Sarpong, E. M. Murray, J. Monsey, J. Zhu, and R. Bose, 2017. Biophysical evidence for intrinsic disorder in the C-terminal tails of the epidermal growth factor receptor (EGFR) and HER3 receptor tyrosine kinases. *Journal of Biological Chemistry* 292:597–610.

32. Wright, P. E., and H. J. Dyson, 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews. Molecular cell biology* 16:18–29.

33. Shoemaker, B. A., J. J. Portman, and P. G. Wolynes, 2000. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proceedings of the National Academy of Sciences* 97:8868–8873.

34. Huang, Y., and Z. Liu, 2009. Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding-Binding Process: A Critical Assessment of the "Fly-Casting" Mechanism. *Journal of Molecular Biology* 393:1143–1159.

35. Zhou, H. X., X. Pang, and C. Lu, 2012. Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets. *Physical Chemistry Chemical Physics* 14:10466–10476.

36. Ruan, H., Q. Sun, W. Zhang, Y. Liu, and L. Lai, 2019. Targeting intrinsically disordered proteins at the edge of chaos. *Drug Discovery Today* 24:217–227.

37. Sibille, N., and P. Bernadó, 2012. Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochemical Society Transactions* 40:955–962.

38. Vranken, W. F., W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue, 2005. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* 59:687–696.

39. Lipari, G., and A. Szabo, 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society* 104:4546–4559.

40. Hall, C. K., and E. Helfand, 1982. Conformational state relaxation in polymers: Time-correlation functions. *The Journal of Chemical Physics* 77:3275–3282.

41. Dejean de la Batie, R., F. Lauprêtre, and L. Monnerie, 1988. Carbon-13 NMR Investigation of Local Dynamics in Bulk Polymers at Temperatures Well Above the Glass Transition. 1. Poly(viny1 methyl ether). *Macromolecules* 21:2045–2052.

42. Dejean de la Batie, R., F. Lauprêtre, and L. Monnerie, 1988. Carbon-13 NMR Investigation of Local Dynamics in Bulk Polymers at Temperatures Well Above the Glass Transition Temperature. 2. Poly(propy1ene oxide) and Linear and Cross-Linked Poly(ethy1ene oxides). *Macromolecules* 21:2052–2058.

43. Bahar, I., B. Erman, and L. Monnerie, 1994. Effect of molecular structure on local chain dynamics: Analytical approaches and computational methods. *Advances in Polymer Science* 116:145–206.

44. Fitzgerald, J. E., A. K. Jha, T. R. Sosnick, and K. F. Freed, 2007. Polypeptide motions are dominated by peptide group oscillations resulting from dihedral angle correlations between nearest neighbors. *Biochemistry* 46:669–682.

45. Avdoshenko, S. M., A. Das, R. Satija, G. A. Papoian, and D. E. Makarov, 2017. Theoretical and computational validation of the Kuhn barrier friction mechanism in unfolded proteins. *Scientific Reports* .

46. Echeverria, I., D. E. Makarov, and G. A. Papoian, 2014. Concerted dihedral rotations give rise to internal friction in unfolded proteins. *Journal of the American Chemical Society* 136:8708–8713.

47. Chou, J. J., S. Gaemers, B. Howder, J. M. Louis, and A. Bax, 2001. A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *Journal of Biomolecular NMR* 21:377–382.

48. Ozenne, V., F. Bauer, L. Salmon, J.-r. Huang, M. R. Jensen, S. Segard, P. Bernado, C. Charavay, and M. Blackledge, 2012. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* 28:1463–1470.

49. Vuister, G. W., and A. Bax, 1993. Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha)) Coupling-Constants in N-15-Enriched Proteins. *Journal of the American Chemical Society* 115:7772–7777.

50. Lescop, E., P. Schanda, and B. Brutscher, 2007. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of Magnetic Resonance* 187:163–169.

51. Lescop, E., R. Rasia, and B. Brutscher, 2008. Hadamard Amino-Acid-Type Edited NMR Experiment for Fast Protein Resonance Assignment. *Journal of the American Chemical Society* 130:5014–5015.

52. Sharp, P., and V. A. Bloomfield, 1968. Light scattering from wormlike chains with excluded volume effects. *Biopolymers* 6:1201–1211.

53. Pedersen, J. S., and P. Schurtenberger, 1996. Scattering functions of semiflexible polymers with and without excluded volume effects. *Macromolecules* 29:7602–7612.

54. Obrien, D. P., B. Hernandez, D. Durand, V. Hourdel, A. C. Sotomayor-Pérez, P. Vachette, M. Ghomi, J. Chamot-Rooke, D. Ladant, S. Brier, and A. Chenal, 2015. Structural models of intrinsically disordered and calcium-bound folded states of a protein adapted for secretion. *Scientific Reports* 5:1–11.

55. Receveur-Brechot, V., and D. Durand, 2012. How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective. *Current Protein & Peptide Science* 13:55–75.

56. Bernadó, P., and M. Blackledge, 2009. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophysical Journal* 97:2839–2845.

57. Wilkins, D. K., S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, and L. J. Smith, 1999. Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques. *Biochemistry* 38:16424–16431.

58. Marsh, J. a., V. K. Singh, Z. Jia, and J. D. Forman-Kay, 2006. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Science* 15:2795–2804.

59. Bernado, P., L. Blanchard, P. Timmins, D. Marion, R. W. H. Ruigrok, and M. Blackledge, 2005. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proceedings of the National Academy of Sciences* 102:17002–17007.

60. Mohana-Borges, R., N. K. Goto, G. J. A. Kroon, H. J. Dyson, and P. E. Wright, 2004. Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *Journal of Molecular Biology* 340:1131–1142.

61. Fuxreiter, M., I. Simon, P. Friedrich, and P. Tompa, 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *Journal of Molecular Biology* 338:1015–1026.

62. Ferreon, J. C., and V. J. Hilser, 2003. The effect of the polyproline II (PPII) conformation on the denatured state entropy. *Protein science : a publication of the Protein Society* 12:447–457.

63. Shen, Y., J. Roche, A. Grishaev, and A. Bax, 2018. Prediction of nearest neighbor effects on backbone torsion angles and NMR scalar coupling constants in disordered proteins. *Protein Science* 27:146–158.

64. Jensen, M. R., G. Communie, E. A. J. Ribeiro, N. Martinez, A. Desfosses, L. Salmon, L. Mollica, F. Gabel, M. Jamin, S. Longhi, R. W. H. Ruigrok, and M. Blackledge, 2011. Intrinsic disorder in measles virus nucleocapsids. *Proceedings of the National Academy of Sciences* 108:9839–9844.

65. Klein-Seetharaman, J., M. Oikawa, S. B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L. J. Smith, C. M. Dobson, and H. Schwalbe, 2002. Long-Range Interactions Within a Nonnative Protein. *Science* 295:1719–1722.

66. Khan, S. N., C. Charlier, R. Augustyniak, N. Salvi, V. Dé Jean, G. Bodenhausen, O. Lequin, P. Pelupessy, and F. Ferrage, 2015. Distribution of Pico- and Nanosecond Motions in Disordered Proteins from Nuclear Spin Relaxation. *Biophysical Journal* 109:988–999.

67. Gast, K., H. Damaschun, K. Eckert, K. Schulze-Forster, H. R. Maurer, M. Mueller-Frohne, D. Zirwer, J. Czarnecki, and G. Damaschun, 1995. Prothymosin .alpha.: A Biologically Active Protein with Random Coil Conformation. *Biochemistry* 34:13211–13218.

68. Schwalbe, H., K. M. Fiebig, M. Buck, J. A. Jones, S. B. Grimshaw, A. Spencer, S. J. Glaser, L. J. Smith, and C. M. Dobson, 1997. Structural and Dynamical Properties of a Denatured Protein. Heteronuclear 3D NMR Experiments and Theoretical Simulations of Lysozyme in 8 M Urea. *Biochemistry* 36:8977–8991.

69. Kovacs, E., R. Das, Q. Wang, T. S. Collier, A. Cantor, Y. Huang, K. Wong, A. Mirza, T. Barros, P. Grob, N. Jura, R. Bose, and J. Kuriyan, 2015. Analysis of the Role of the C-Terminal Tail in the Regulation of the Epidermal Growth Factor Receptor. *Molecular and Cellular Biology* 35:3083–3102.

70. Wright, P. E., and H. J. Dyson, 2009. Linking folding and binding. *Current Opinion in Structural Biology* 19:31–38.

71. Arai, M., K. Sugase, H. J. Dyson, and P. E. Wright, 2015. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proceedings of the National Academy of Sciences* 112:9614–9619.

72. Oldfield, C. J., Y. Cheng, M. S. Cortese, P. Romero, V. N. Uversky, and A. K. Dunker, 2005. Coupled Folding and Binding with α-Helix-Forming Molecular Recognition Elements. *Biochemistry* 44:12454–12470.

73. Zhang, X., J. Gureasko, K. Shen, P. A. Cole, and J. Kuriyan, 2006. An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell* 125:1137–1149.

74. Pines, G., P. H. Huang, Y. Zwang, F. M. White, and Y. Yarden, 2010. EGFRvIV: A previously uncharacterized oncogenic mutant reveals a kinase autoinhibitory mechanism. *Oncogene* 29:5850–5860.

75. Sorkin, A., M. Mazzotti, T. Sorkina, L. Scotto, and L. Beguinot, 1996. Epidermal Growth Factor Receptor Interaction with Clathrin Adaptors Is Mediated by the Tyr 974 -containing Internalization Motif. *Journal of Biological Chemistry* 271:13377–13384.

76. Iešmantavičius, V., M. R. Jensen, V. Ozenne, M. Blackledge, F. M. Poulsen, and M. Kjaergaard, 2013. Modulation of the intrinsic helix propensity of an intrinsically disordered protein reveals long-range helix-helix interactions. *Journal of the American Chemical Society* 135:10155–10163.

77. Cadena, D. L., C. L. Chan, and G. N. Gill, 1994. The intracellular tyrosine kinase domain of the epidermal growth factor receptor undergoes a conformational change upon autophosphorylation. *Journal of Biological Chemistry* 269:260–265.

78. Mansiaux, Y., A. P. Joseph, J. C. Gelly, and A. G. de Brevern, 2011. Assignment of polyproline ii conformation and analysis of sequence - structure relationship. *PLoS ONE* 6:1–15.

79. Pawson, T., 1992. SH2 and SH3 domains. *Current Opinion in Structural Biology* 2:432–437.

80. Macias, M. J., S. Wiesner, and M. Sudol, 2002. WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Letters* 513:30–37.

81. Saksela, K., and P. Permi, 2012. SH3 domain ligand binding: What's the consensus and where's the specificity? *FEBS Letters* 586:2609–2614.

82. Bornet, O., M. Nouailler, M. Feracci, C. Sebban-Kreuzer, D. Byrne, H. Halimi, X. Morelli, A. Badache, and F. Guerlesquin, 2014. Identification of a Src kinase SH3 binding site in the C-terminal domain of the human ErbB2 receptor tyrosine kinase. *FEBS Letters* 588:2031–2036.

83. Xie, Y., A. M. Pendergast, and M. C. Hung, 1995. Dominant-negative mutants of Grb2 induced reversal of the transformed phenotypes caused by the point mutation-activated rat HER-2/Neu. *The Journal of biological chemistry* 270:30717–24.

84. Wulf, G., P. Garg, Y.-C. Liou, D. Iglehart, and K. P. Lu, 2004. Modeling breast cancer in vivo and ex vivo reveals an essential role of Pin1 in tumorigenesis. *The EMBO Journal* 23:3397–3407.

85. Chen, C., Z. Zhou, R. Liu, Y. Li, P. B. Azmi, and A. K. Seth, 2008. The WW domain containing E3 ubiquitin protein ligase 1 upregulates ErbB2 and EGFR through RING finger protein 11. *Oncogene* 27:6845–6855.

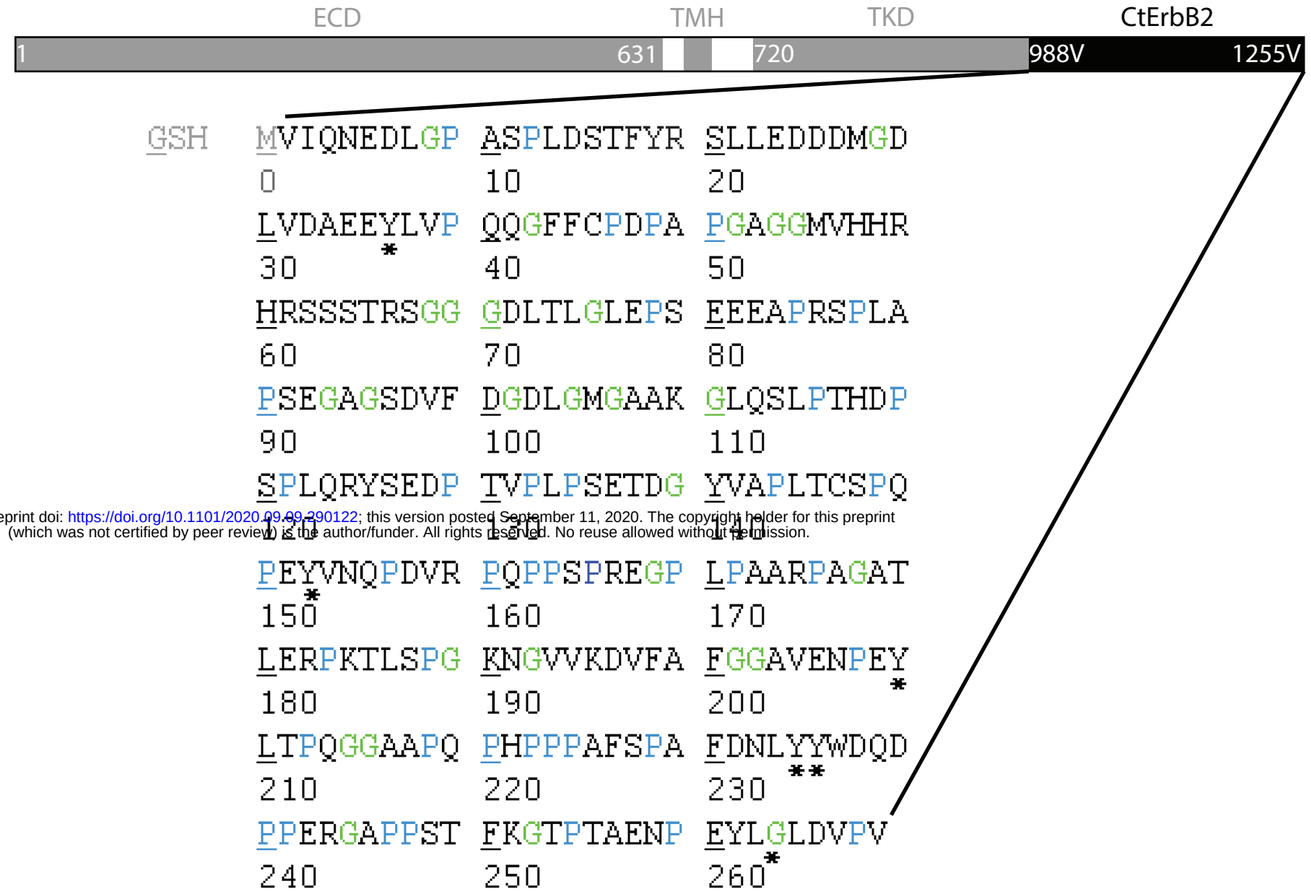
86. Khurshid, R., M. Saleem, G. e Raana, and M. S. Akhthar, 2014. Phosphorylation sites of HER2/c-erbB-2: role in cell growth and in disease. *Acta biochimica Polonica* 61:699–703.

87. Rice, P., L. Longden, and A. Bleasby, 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* 16:276–277.

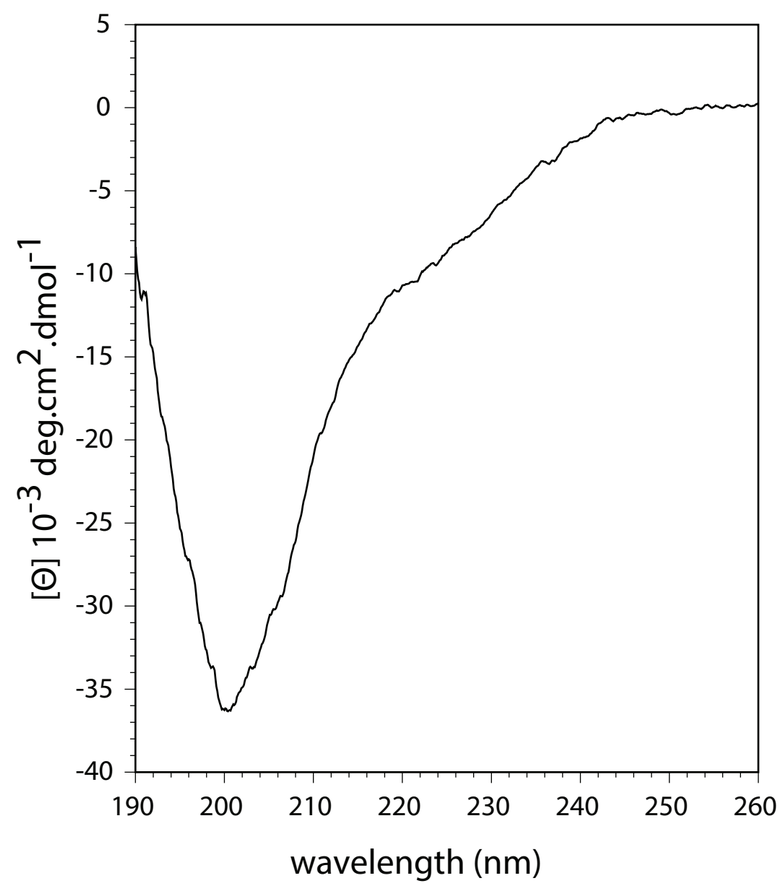
SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

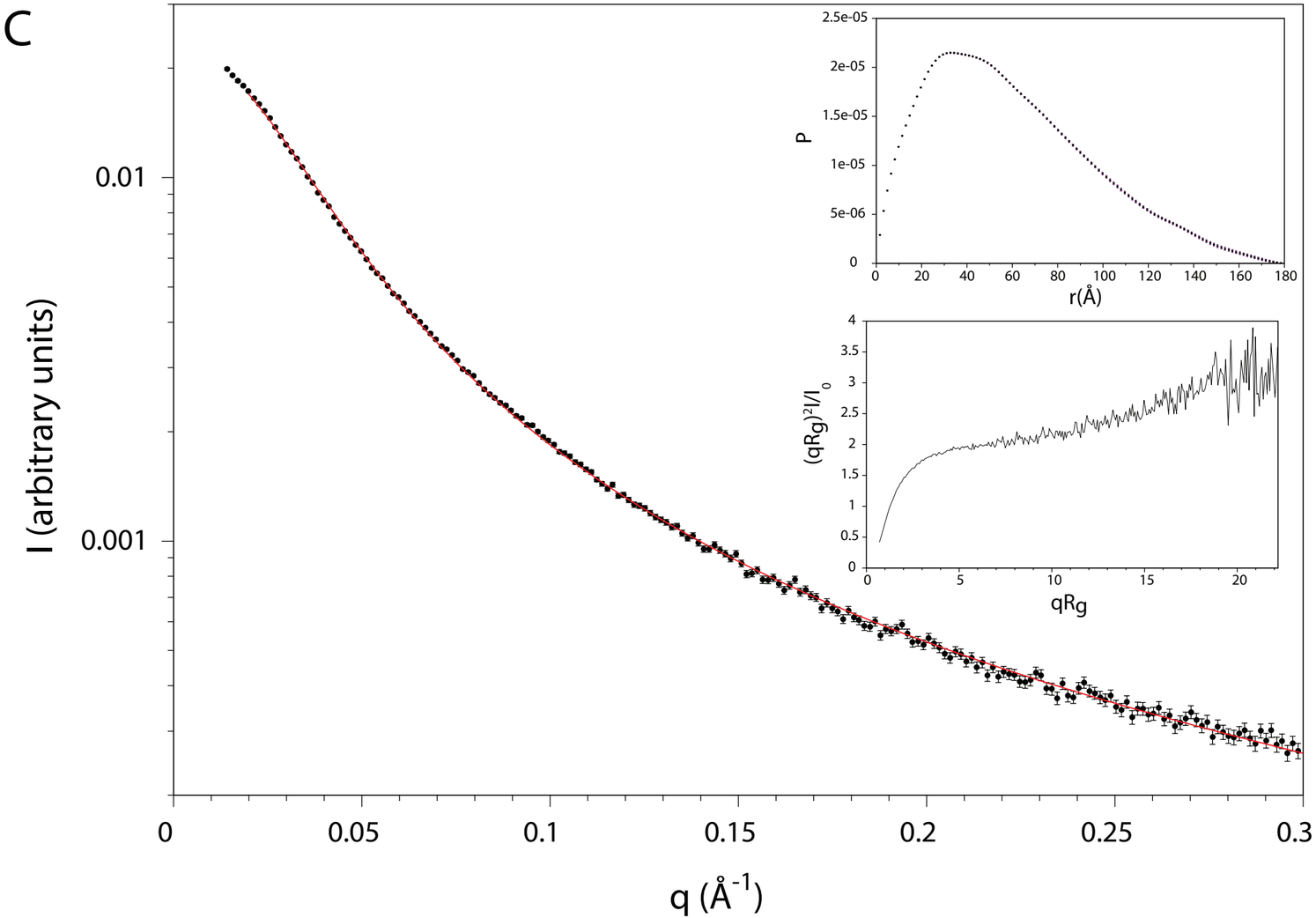
A



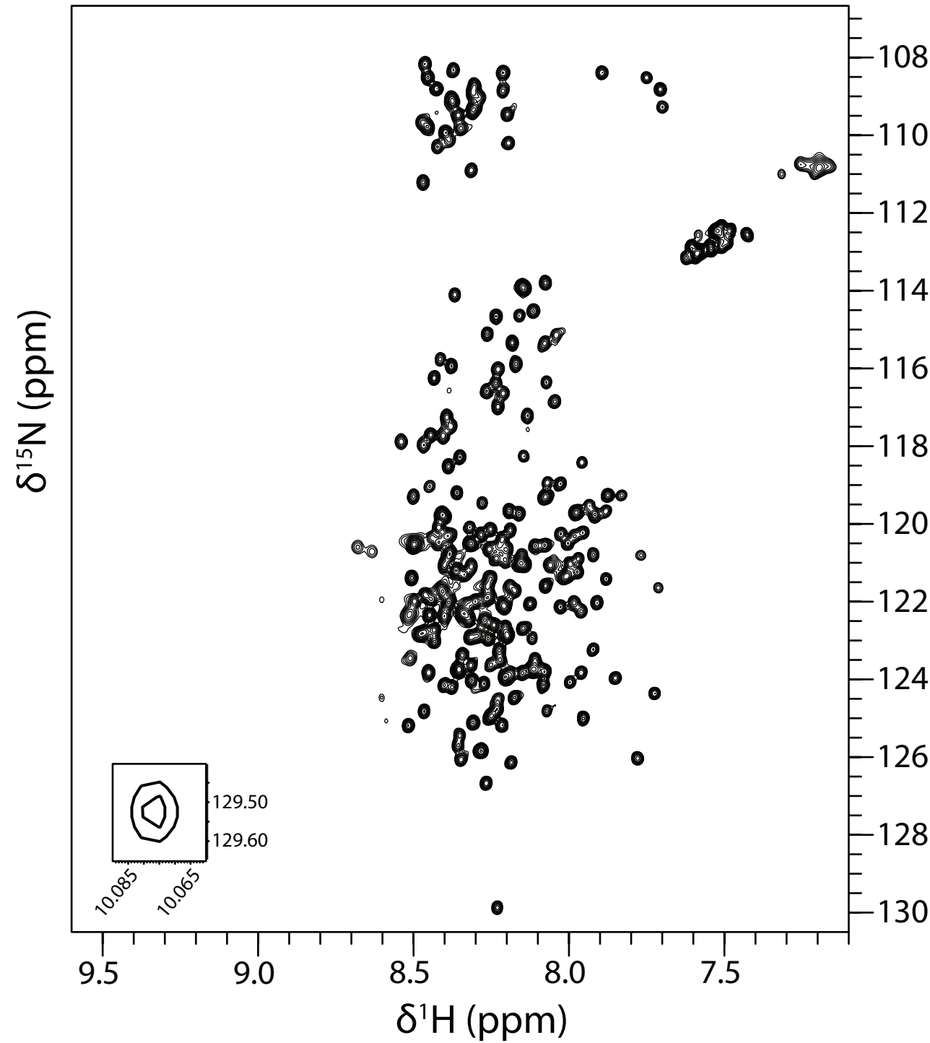
B

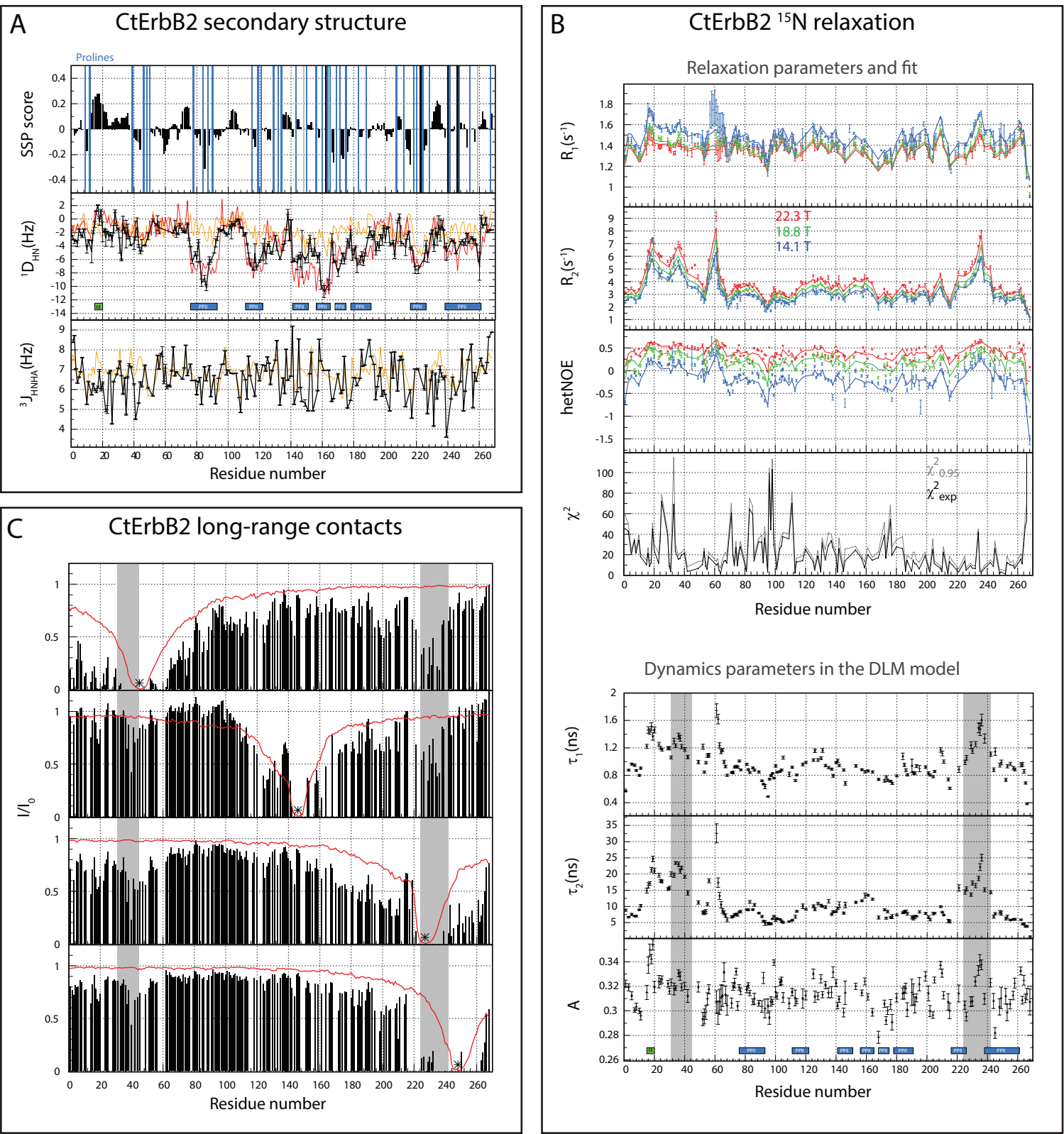


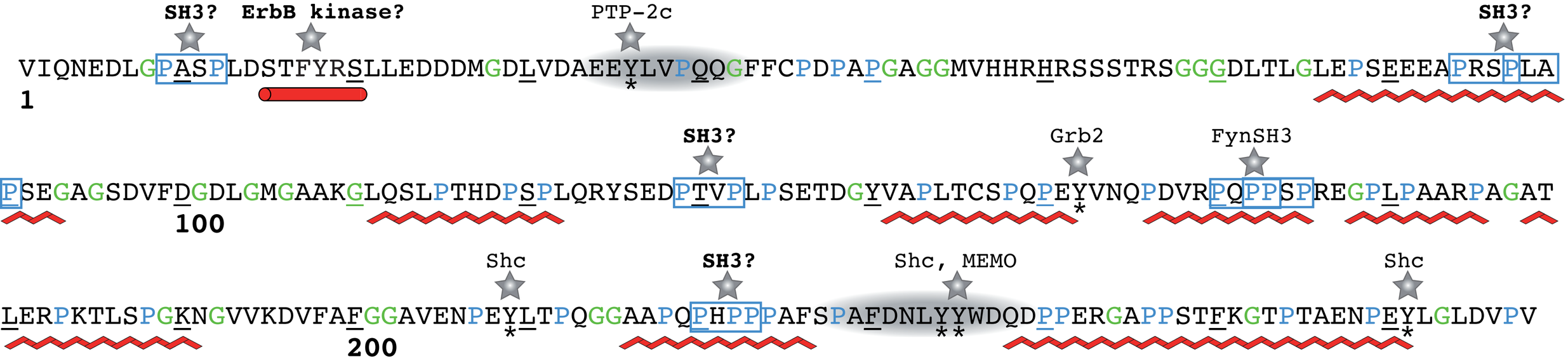
C



D




A
CtErbB2 secondary structure
Prolines
SSP score
$^1D_{HN}$(Hz)
$^3J_{HNHA}$(Hz)
Residue number
B
CtErbB2 ^{15}N relaxation
Relaxation parameters and fit
$R_1(s^{-1})$
$R_2(s^{-1})$
22.3 T
18.8 T
14.1 T
hetNOE
χ^2
$\chi^2_{0.95}$
χ^2_{exp}
Residue number
Dynamics parameters in the DLM model
τ_1(ns)
τ_2(ns)
A
Residue number
C
CtErbB2 long-range contacts
I/I$_0$
Residue number



PPII helix

α helix

main known or putative interaction site

PxxP motif

zone of long-range contact

* tyrosine autophosphorylation site