HAL
open science

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, Dominique Durand, N. Assrir, et al.

To cite this version:

L. Pinet, Y.-H. Wang, C. Deville, E. Lescop, F. Guerlesquin, et al.. Structural and dynamic characterization of the C-terminal tail of ErbB2: disordered but not random. Biophysical Journal, 2021, 120 (10), pp.1869-1882. 10.1016/j.bpj.2021.03.005 . hal-03095358v2

HAL Id: hal-03095358

https://hal.science/hal-03095358v2

Submitted on 26 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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]Institut de Chimie des Substances Naturelles (ICSN), CNRS UPR2301, Université Paris-Saclay, Gif-sur-Yvette, France
[2]Department of Biochemistry, University of Zurich, Zurich, Switzerland
[3]SGS Taiwan LTD, No.38, Wu Chyuan 7th Rd., New Taipei Industrial Park, Wu Ku District, New Taipei City, 24890 Taiwan
[4]IGBMC, University of Strasbourg, CNRS UMR, Illkirch, 7104, France
[5]LISM, Institut de Microbiologie de la Méditerranée, CNRS and Aix-Marseille University, Marseille, France
[6]Centre de Recherche en Cancérologie de Marseille (CRCM), Aix-Marseille Univ, INSERM, Institut Paoli-Calmettes, CNRS, Marseille, France
[7]Department of Virology, Institut Pasteur, CNRS UMR 3569, Paris, France
[8]I2BC, Université Paris-Saclay, CNRS UMR 9198, Gif-sur-Yvette, France
[*]Correspondence: carine.van-heijenoort@cnrs.fr

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, complemented by SAXS. We show that although no part of CtErbB2 has any fully populated secondary or tertiary structure, it contains several transient α helices and numerous transient PPII helices, populated up to 20 and 40% respectively, and low but significant compaction. The presence of some structural elements suggests, along the lines of the results obtained for EGFR (ErbB1), that they may have a functional role in ErbB2's auto-regulation processes. In addition, the transient formation of PPII helices is compliant with previously suggested interactions with SH3 domains. All in all, our in-depth structural study opens new perspectives in the mechanistic understanding of ErbB2.

SIGNIFICANCE We report here the first description of the receptor tyrosine kinase ErbB2 disordered tail (CtErbB2) at atomic resolution using NMR, complemented by SAXS. We show that although CtErbB2 is an intrinsically disordered region, it exhibits significantly populated local secondary structure elements and transient long-range contacts. As it has been previously shown for EGFR, some of these structural features could have a role in the regulation of the receptor's activity. In addition, the evidence of several significantly populated PPII helices supports the possibility of CtErbB2 interactions with SH3 protein domains at sites distinct from the main phosphotyrosine 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 (3). The kinase domains are activated in the dimers (or potential higher-order oligomers (4)), leading to tyrosine phosphorylation of the C-terminal tails (5). 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 (5).

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 (6). ErbB2 upregulation is found in up to 20% of breast cancers and correlates with poor prognosis (7), 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 (8, 9). These strategies have been widely successful, but are limited by inherent or acquired resistance, calling for new combined therapeutic approaches and dual HER2-targeting (10) that require an extended knowledge of associated molecular mechanisms. While elucidating the structures of the extracellular (11–19), transmembrane (20–22), juxtamembrane (22) and kinase (23, 24) 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% (25). Each phosphotyrosine interacts with distinct partners, SH2 and PTB (phosphotyrosine-binding) domains-containing proteins or MEMO (mediator of ErbB2-driven cell motility) (5, 26, 27), to trigger different signaling pathways *in vivo* (25, 28). 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 (29, 30) and regulates kinase activity (31–33). Similar mechanisms could be at stake in CtErbB2 function.

We previously performed the NMR assignment of CtErbB2 and demonstrated that it is intrinsically disordered (34), in agreement with predictions based on its sequence composition and with measurements of solvent accessibility (35). Intrinsically disordered proteins (IDPs) are a class of proteins with no major, stable three-dimensional 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 (36), most likely due to their specific mechanical properties, their increased capture radius ("fly-casting" mechanism (37, 38)) and their specific but usually low- to medium-affinity binding (39). 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 (40). 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) (41). Here we present the first atomic-scale description of CtErbB2 conformational ensemble, using both NMR and SAXS. We show that although this region is highly disordered, it exhibits partially formed α helices, and transient long-range contacts. Several proline-rich regions also transiently adopt polyproline II (PPII) helices including potential SH3- and WW-binding short linear motifs. Together, our data show a high level of structural diversity for an intrinsically disordered region. The possible functional relevance of these structural and dynamic characteristics, which remains to be studied, is discussed, particularly in the light of those already demonstrated for the EGFR.

MATERIALS AND METHODS

Protein expression and purification

CtErbB2 was expressed and purified as described in our previous study (34). The final construct, depicted in Fig.1 A, comprises residues 988-1255 of the full-length human ErbB2 receptor (which were numbered from 1 to 268 in our assignment publication (34)), plus four additional N-terminal residues (Gly-Ser-His-Met, numbered from -4 to -1).

NMR spectroscopy

The typical samples for NMR studies were composed of 200-400 μM protein, in 200-250 μL of 2-(N-morpholino)ethanesulfonic acid (MES) buffer pH 5.6 (40 mM MES, 200 mM NaCl, 2 mM tris(2-carboxyethyl)phosphine (TCEP), 5% D_2O). At pH closer to 7, a lot of peaks experienced reduced intensity (see Fig. S1 in the Supporting Material), due to the increased exchange rate of amide protons with water. NMR spectra were recorded at 298 K as a compromise between stability of the protein, quality of the spectra and physiological relevance. Changes in temperature did not have dramatic effects between 278 and 298 K, as shown in Fig. S2. 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 (42).

^{15}N relaxation ^{15}N relaxation parameters R_1, R_2 and ^{15}N-{^{1}H} 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. A longitudinal relaxation delay (d_1) of 3 s (R_1 and R_2) or 4 s (nOes) was used between each scan. 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-{^{1}H} nOes were measured using a TROSY-based pulse sequence using 120° pulses every 5 ms for 4 s for proton saturation. The ^{15}N-{^{1}H} 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 (43) and from the Hall and Helfand polymer model (44). 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 (45, 46) (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 (47, 48). 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 $(\phi_{i-1}, \psi_i) \leftrightarrow (\phi_i, \psi_{i+1})$ ((49, 50). 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.* (51). 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 (52) on two ensembles of 10 000 structures generated without constraints (random coil) or with the following secondary structure propensities: an α helix at 20% between residues 1002-1007, PPII helices (defined with dihedral angles (-78°, 149°)) at 20% (residues 1063-1080, 1165-1178, 1203-1213 and 1225-1248), 30% (residues 1098-1109, 1128-1138 and 1155-1162) and 40% (residues 1143-1152).

Scalar couplings measurements The homonuclear $^3J_{HNHA}$ coupling constant for each residue was measured with the 3D HNHA experiment of Vuister and Bax (53) implemented by Bruker. Data analysis to extract coupling constant from the intensity ratio of the cross- and diagonal-peaks was performed directly in the dedicated section of CCPNMR, also allowing to extract error on the constant using the noise level of the spectrum.

Paramagnetic Relaxation Enhancement Four mutants of CtErbB2 were used for four different experiments: C1133S, C1032S, (C1032S,C1133S,S1214C) and (C1032S,C1133S,S1235C) (for respective paramagnetic tag positions of C1032, C1133, C1214 and C1235). These ^{15}N-^{13}C labeled mutants were reduced with TCEP, buffer-exchanged to 20 mM sodium borate buffer at pH 8.0 without TCEP (ZebaSpin Desalting columns, ThermoScientific), and incubated at 15°C overnight with the paramagnetic probe (1:10 protein:probe ratio). The buffer was then exchanged for the MES buffer pH 5.6, removing the excess of free paramagnetic probe. The BEST-^{15}N-^{1}H 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 BEST-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 (54) and HADAMAC (55) experiments.
The probe that was used in this study is 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 unlabeled protein for all mutants, indicating no major perturbation of the conformational ensemble.

Small angle X-ray scattering

In a preliminary experiment we recorded several SAXS curves on a Nanostar Instrument (Bruker) with a Microstar rotating anode (λ = 1.54 Å) at 285 K, at concentrations from 50 to 200 μM. The curves were identical (Fig. S3), indicating the absence of oligomerization effects.
Main SAXS measurements were carried out on the SWING beamline of the SOLEIL Synchrotron Radiation Facility (Saint-Aubin, France). The sample to detector (Eiger) distance was set to 2000 mm, allowing reliable data collection over the momentum transfer range 0.005 Å^{-1} $< q <$ 0.5 Å^{-1} with $q = 4\pi sin(\theta)/\lambda$ where 2θ is the scattering angle and λ the wavelength of the X-rays (λ = 0.1 nm). To use mono-disperse solutions devoid of aggregates, SAXS data were collected directly after elution through a size-exclusion high-performance liquid chromatography (SEC-HPLC) column (Agilent Bio SEC-3) mounted on-line with the SAXS flowthrough capillary cell (1.5 mm in diameter) under vacuum. Sample concentration was 190 μM before injection on the SEC column. The protein was eluted at a flow rate of 300 μL/min. Temperature of the SAXS measurements in the SAXS capillary and in the column during elution was 298 K. During elution scattering patterns (frames) were recorded for 990 ms with a dead time between frames of 10 ms. Twenty identical frames corresponding to the top of the elution peak were averaged. SAXS data was normalized to the intensity of the incident beam and background (i.e., the elution buffer) subtracted using the programs FoxTrot (courtesy of SWING beamline) and Primus. The scattered intensities were plotted on an absolute scale using water scattering.
A Sharp-Bloomfield (56, 57) equation was used to fit the data and obtain polymer parameters for the polypeptide chain (58, 59):

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

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 $L = Nl_0 f$ 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 (59).

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 (CD) of CtErbB2 was measured at 298 K and shows the characteristic signature of an IDP, with only one strong negative band at 200 nm (60) (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 53.6 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; $\nu = \frac{1}{3}$ in poor solvent conditions (globular proteins), $\nu = \frac{1}{2}$ in a θ solvent (pure random walk), and $\nu = \frac{3}{5}$ in a good solvent (expanded coil, with an excluded volume). Its applicability to denatured or disordered proteins has already been investigated (59). Bernadó and Blackledge (61) found that parameters R_0 = 2.54 Å and ν = 0.522 recapitulated well the behavior of such chains. For CtErbB2, this equation gives R_g = 47.4 Å, relatively close to 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 ν = 0.29 (62).
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 (34) and the chemical shifts were deposited into the BMRB (entry 26740). $\delta(^{13}C_\alpha)$, $\delta(^{13}C_\beta)$ and $\delta(^1H_\alpha)$ were used as input for the SSP algorithm (63), and the results are reproduced 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 ($^1D_{NH}$, Fig.2 A) are negative almost all along CtErbB2 sequence, as usually observed for denatured or disordered proteins (64, 65). Overall, all the data show that CtErbB2 is an intrinsically disordered region (IDR).

CtErbB2 displays local propensities for a residual α helix and numerous residual 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 (66). 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 ($^3J_{HNHA}$) and $^1D_{NH}$ 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 (67). These left-handed

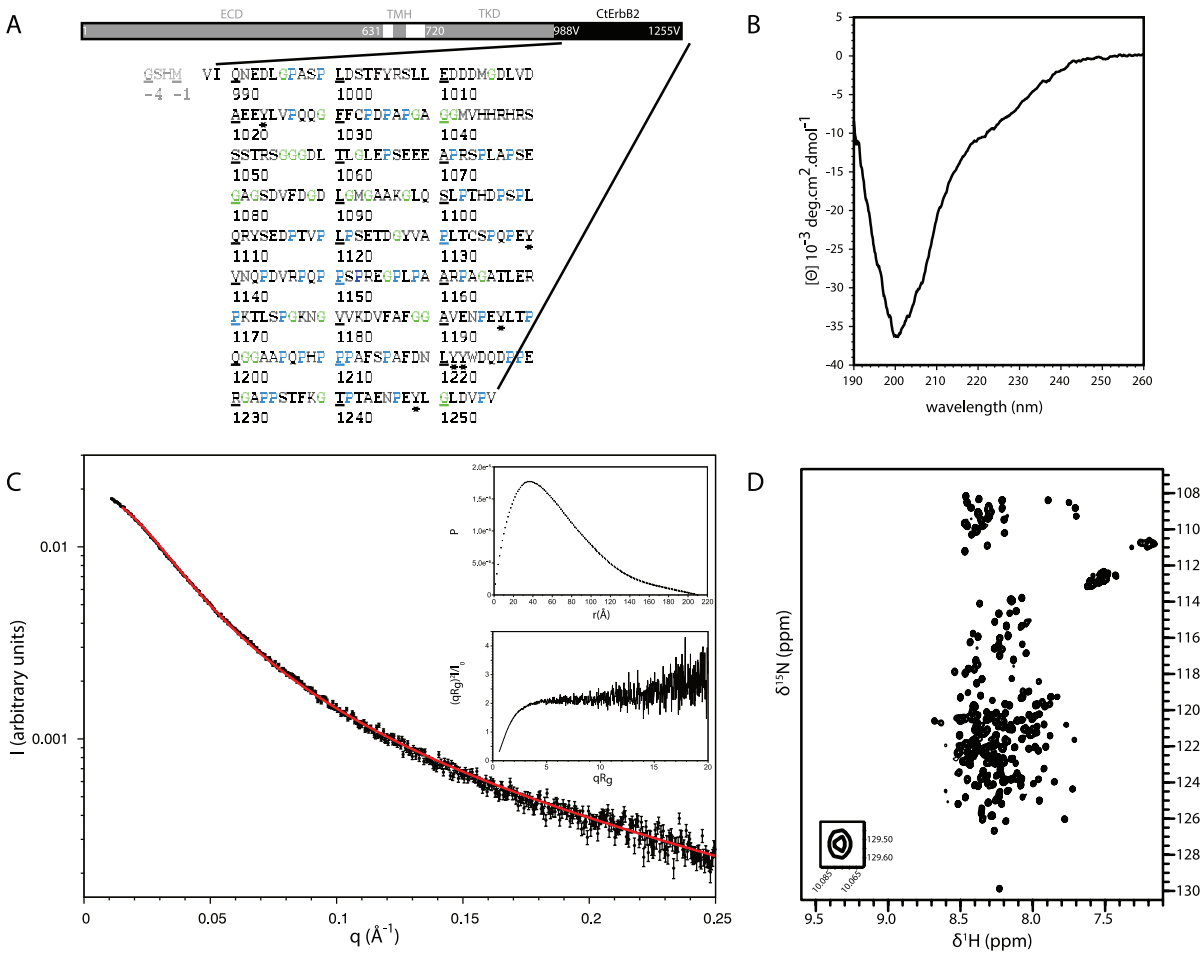

Figure 1: ***(A)*** Schematic representation of the human ErbB2 receptor domains and sequence of the CtErbB2 construct. ECD = extracellular domain; TMH = transmembrane helix; TKD = tyrosine kinase domain; CtErbB2 = C-terminal tail. Our CtErbB2 construct starts at valine 988 in the full length receptor, with four additional N-terminal residues. 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.

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 (34) 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 (Fig. S4). However, due to still insufficient signal-to-noise ratio, no sequential assignment of these peaks could be achieved. The ratio of average peak volumes corresponding to residues following *cis* and *trans* prolines gives an estimate of 8% for the *cis*-proline conformation.

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 (63).

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 (68). Indeed, residues involved in β strands have higher $^{3}J_{HNHA}$ (more than 8 Hz) than those involved in PPII helices, which exhibit $^{3}J_{HNHA}$ around 6 Hz, close

to or below the random coil value (around 7 Hz) (68, 69). In Fig.2 A are presented $^3J_{HNHA}$ values, compared to random coil $^3J_{HNHA}$ predicted for this sequence (https://spin.niddk.nih.gov/bax/nmrserver/rc_3Jhnha/ (70)). Experimental $^3J_{HNHA}$ being almost always equal or below the predicted random coil value, no propensity for β strands can be detected in CtErbB2, in 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 α helices, that are all N-capped by a serine or aspartate as commonly observed (71) (^{1001}DSTFYRSLL^{1009}, ^{1058}DLTLGL^{1063}, ^{1087}DGDLGMG^{1093}, ^{1122}SETDGY^{1127}, ^{1218}DNLYY^{1222}). 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 (72). CtErbB2 N-terminal helix is longer and more populated (around 20% according to SSP) than the following ones. It corresponds to high $^1D_{NH}$, consistent with N-H bonds of the α-helix partially aligned parallel to the gel stretching direction and the magnetic field, and low $^3J_{HNHA}$ values (see Fig.2 A). With their short length and lower population, the other helices do not seem to significantly affect the conformational ensemble (no major deviation in J and dipolar couplings compared to the random coil values in orange). They can still be functionally relevant, but were not taken into account here for the generation of a conformational ensemble.

In conclusion, 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 and 40%. The $^1D_{NH}$ RDCs back-calculated from 10000 conformers generated in the software Flexible Meccano (52) with no secondary structure propensities significantly deviate from the experimental values, whereas those obtained using these propensities as input constraints correctly account for the experimental values (orange and red lines in Fig.2 A, respectively).

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 bonds 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 (37) 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 high microsecond to millisecond timescale.
^{15}N R_2 and R_1 and {^{1}H}-^{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 panel 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 (73). To gain more insight into CtErbB2 dynamics behavior, the data were first analyzed using the classically used "model-free" approaches, which did not yield statistically relevant fits. Data could be properly fitted with the polymer model developed by DeJean de la Batie, Lauprêtre and Monnerie (45, 46), 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): A, 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 (74).
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 bonds 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 (1052-1097), 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 ($G_{1055}G_{1056}G_{1057}, G_{1188}G_{1189}, G_{1201}G_{1202}$), or glycine rich segment (1080-1097) with 6 glycines.

- (ii) The histidine rich region (1043-1048) exhibits much more restricted motions with highest τ_1 and τ_2 (almost 2 and

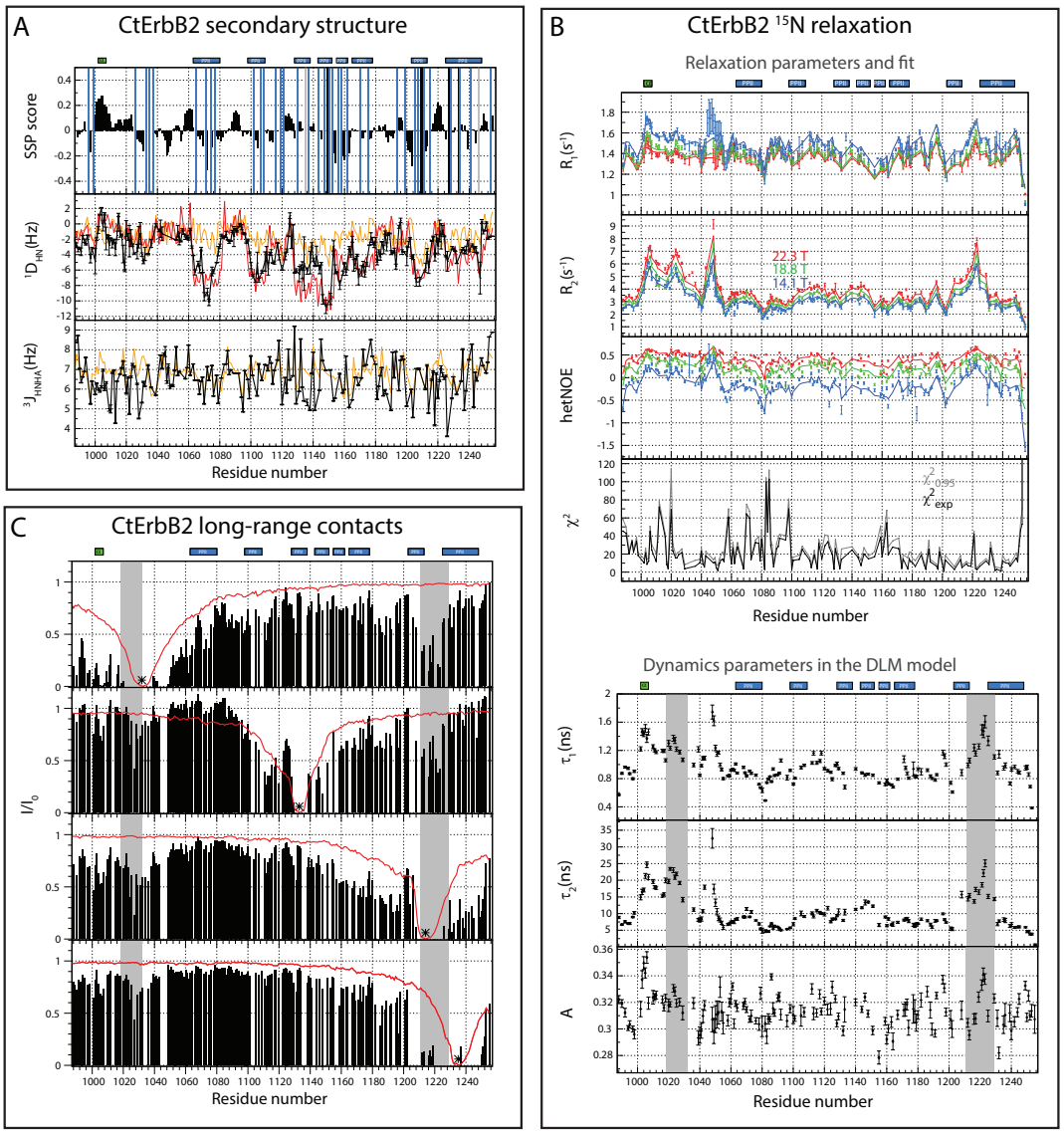

Figure 2: CtErbB2 secondary structure and long-range contacts identification. Green box= α helix; blue box = PPII helix. Grey area: zones of tertiary contact. (***A***) *Top*: SSP scores (63), reproduced from (34). Vertical bars represent the position of prolines. Blue bars: *trans* prolines (with small (<10%) or no *cis* conformation detected); Grey bars: prolines with a major population of *cis* isomer detected; Black bars: prolines of unknown conformation. *Middle*: N-H residual dipolar couplings compared with Flexible Meccano (52) 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 transient secondary structures presented above. *Bottom*: $^3J_{HNHA}$. Black: experimental; orange: calculated for a random coil (70). (***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 panel: 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 C1133S, C1032S, (C1032S,C1133S,S1214C) and (C1032S,C1133S,S1235C) of CtErbB2. The paramagnetic probe, indicated by an asterisk, was attached to cysteine 1032, 1133, 1214 and 1235, respectively. The red line is the PRE values predicted for the ensemble generated by Flexible Meccano taking into account secondary structures.

35 ns respectively). Hydrogen bonds can hinder the flexibility of this highly polar stretch, which also contains several arginines and serines.

- (iii) The two segments (1002-1036) and (1212-1231) 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 (1002-1007) segment, with also slightly restricted libration (increased value of A). The other two regions (1018-1036) and (1212-1231) display only small helical propensities, and therefore their peculiar dynamic features can not be explained only with secondary structure considerations. Transient long-range contacts may be at stake here, that could impede the damping processes of the bonds 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 (73, 75, 76), 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 structure and compaction of a disordered protein (41). 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 (58). 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 = 25.3 Å (which gives a persistence length of 12.65 Å), the contour length L = 653 Å and the radius of gyration of cross section (thickness of the chain) R_C = 4.2 Å. The 12.65 Å persistence length corresponds to a seven to eight amino-acids statistical chain, close to the value of seven estimated by Schwalbe et al.(76) 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 may have an effect, via reduced C_α- C_α distances. The effect of structure formation on the radius of gyration is more complex, since R_g can be modified by both secondary structure and global compaction. Comparison of experimental R_g with those of simulated conformational ensembles is a useful tool to assess the different contributions. Here, the ensemble generated by Flexible Meccano with secondary structures has an average R_g of 57.0 Å, slightly higher than the measured one. The lower experimental R_g could be due to 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 long-range contacts

To investigate the existence and location of long-range 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, C1133S, C1032S, (C1032S, C1133S, S1214C) and (C1032S, C1133S, S1235C), each permitting coupling of MTSL on a single cysteine (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 1214 compared to 1235. This contact is consistent with the increased R_2 values observed in the terminal regions, as highlighted in grey in Fig.2 B. To further confirm that this contact was not induced by an interaction between the MTSL probe and the protein, we verified that the SAXS curves of CtErbB2 with and without MTSL (reduced) on residue C1032 were identical (Fig. S5). Other contacts are observed between the central part and the extremities of CtErbB2 when the probe is attached at position 1133, even though the intensity of the PRE effect is less pronounced in those cases.

Globally, these experiments thus strongly suggest transient long-range contacts, at least between the two segments around 1017-1027 and 1207-1227 of CtErbB2.

Conservation of structural elements of CtErbB2 amongst mammals

We sought to link our observations on structure and dynamics of CtErbB2 to possible 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. S6-S9). 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 conservation score higher than 0.5 as calculated by Clustal Omega). The same observation was made for EGFR (77). Interestingly, the first 50 residues are particularly well conserved (31 are strictly conserved, 10 have scores > 0.5). This region contains the longest N-terminal α helix transiently formed, 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 ^{1215}PAFDNLYYW^{1223} stretch is strictly conserved) and also contains a short transient α helix. 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 might be 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 little sign of the presence of stable secondary structure in the protein. The small positive peak expected for PPII at 220 nm (60) is not seen here, probably due to the transient nature of these structures, and maybe to some compensation from the negative 220nm peak expected for the transient α helices. Deviation from random coil chemical shifts are consistent with many but transient local secondary structures, and relaxation studies show that the whole chain is still very dynamic. Some evidence suggests 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 in the context of the entire protein is the kinase domain, but Keppel *et al.* showed by hydrogen-deuterium exchange followed by mass spectrometry that CtErbB2 is still globally highly solvent exposed in the context of the whole cytoplasmic region, except for its N-terminal extremity adjacent to the kinase (see below) (35). 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. The longest and most populated helix in CtErbB2 is located between residues 1001 and 1009 (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.* (35), 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 (24) and 3PP0 (23)). 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 (32, 77, 78) and in terms of transformation potency (79). A tyrosine phosphorylation was suggested to disrupt this inhibitory intramolecular interaction (32). The EGFR helix involved is called the AP-2 helix, for the FYRAL motif has been shown to interact with the clathrin adaptor protein complex 2 (80). 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. S9). This could support the hypothesis of a similar kinase-inhibitory role for these residues of CtErbB2, but that hypothesis needs biochemical confirmation.

CtErbB2 has long-range contacts involving functionally important regions

The presence of at least one long-range contact is suggested by the slightly smaller R_g derived from SAXS compared to an ensemble with the determined secondary structure propensities. PRE experiments confirmed the presence of such contacts, especially between the N- and C-termini of the tail. Relaxation data confirm that these regions have a restricted 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 around residues 1020 and 1220 (Fig. 2 A). It is not unusual in disordered proteins to observe coincidental secondary and tertiary folding (81). Whether this contact has functional importance remains to be investigated, but its overlap with important sites for signaling is to be noticed. Amongst them are several phosphorylation and interaction sites, such as tyrosine 1023 (the phosphorylation of which was shown to be inhibitory of ErbB2-mediated transformation potency) and tyrosines 1221-1222 (interaction site for Shc) (25). 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 (82), correlating with increased dynamics of the C-terminus (32) and disruption of its interaction with the kinase domain (31). 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(67). It was estimated that about 5% of residues (not only prolines) adopt this left-handed, extended triangular helical conformation of 3 residues per turn (83). 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 (84), and WW domains are sometimes considered as the phosphorylation-dependent equivalent of SH3 domains (85). 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 (86). 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)(85).

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 $^1D_{NH}$ and $^3J_{HNHA}$ in *trans*-proline rich regions shows that the PPII conformation is indeed sampled at multiple positions along the CtErbB2 sequence with populations up to 40%. Three PPII stretches contain PxxP motifs: ^{1068}EEA**P**RS**P**LA**P**SEGA^{1081} (2 sequential motifs sharing one proline), ^{1143}PDVR**P**Q**PP**S**P**REGP^{1156} (two nested motifs, RxxPxxP and PxxPxxR), and ^{1203}AAPQ**P**HP**P**PAFS^{1214}. The first stretch is highly conserved in mammals (stretch 2-3 in Fig. S6). For the second stretch (number 5-6 in Fig. S6) , the 2 nested motifs are not fully conserved, but another PXXP or PXXPXXP motif, with prolines at position 1156, 1159 and 1162, is frequently present in the same PPII rich regions in non human mammalian sequences. Two other PxxP motifs are found: ^{9}92EDLG**P**AS**P**LDST^{1003}, that is highly conserved among mammals (number 1 in Fig. S6), and ^{1112}YSED**P**TV**P**LPSE^{1123}. This last motif is in most non-human sequences part of a PXXPXXP motif with the last proline at position 1122 (stretch 4 in Fig. S6).

The presence of these PPII helices strengthen previous evidence showing binding of ErbB2 to SH3 domains. The isolated peptide ^{1146}RPQPPSPRE^{1154} studied by Bornet *et al.* (87) 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 (88). Indeed, Grb2 SH2 domain (which is known to anchor Grb2 to CtErbB2 via tyrosine 1139) 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 998-999, 1073-1074, 1107-1108, 1134-1135, 1151-1152, 1174-1175 and 1214-1215), and two TP motifs (residues 1198-1199 and 1240-1241). 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(89, 90).

CtErbB2 is known to ensure signal transduction through interactions between its phosphorylated tyrosines and SH2/PTB/MEMO domains (5, 26, 27). PPII structures and proline-containing motifs expand the number of possible molecular recognition features

of CtErbB2, with potential recognition by SH3 or WW domains. On the contrary, no β strand is observed in CtErbB2, although the last few C-terminal residues (peptide EYLGLDVPV) were shown to adopt this structure when bound to ERBIN (PDB 1MFL) (91). Given the great versatility of the modes of binding of IDP recognition motifs, partner recognition mechanisms with possible folding upon binding in CtERbB2 remain to be fully investigated.

CtErbB2 contains two prolines with significant *cis* population

Notable exceptions among prolines are P1135 and P1246, 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 1135 (^{1133}CS**P**QPEYVN^{1141}) and 1246 (^{1244}EN**P**EYLG^{1250})), the *cis* prolines are close to tyrosines 1139 and 1248 that were shown to be phosphorylated and important for function (25, 28, 92). Additionally, the phospho-SP sequence of proline 1135 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. Even though the proximity of aromatic residues can increase the *cis* population (93), no sequence or structure determinant can explain the presence of these populated *cis* prolines at these specific positions and not others in the sequence.

CONCLUSION

The function of intrinsically disordered domains involved in signaling processes such as CtErbB2, is related both to their degree of global disorder and to the presence along their sequence of accessible key motifs likely to interact with specific partners, due to their mere physico-chemical characteristic and/or their ability to locally adopt a specific spatial organization. The structural and dynamic features of CtErbB2 that we have highlighted in this work may in this line be related to several reported functions of the disordered tail in the receptor : the global disorder of the chain ensures high accessibility and adaptability for post-translational modifications and multiple interaction with partners; the presence of long-range contacts, although transient, can modify the conformational ensemble; the significant presence of an α helix similar to one identified in EGFR could participate in the same way to kinase activity regulation; the identification of significant amount of PPII helices enable possible interactions with SH3 or WW protein domains. These structural and potential functional features are summarized in Fig. 3, together with parts of the conformational ensemble generated by Flexible Meccano taking into account secondary structures propensities. Our results constitute a first step to be able to now dissect the functional role of the structural elements of CtErbB2, and to move on to other functionally relevant conditions, such as the implication of tyrosine phosphorylation on these structural features and their impact on the protein interaction network and the regulation of signal transduction.

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. We also thank Arthur Besle and Anaïs Vogel for their help in protein preparation and Annie Moretto for buffer, reagents and media preparation. We acknowledge SOLEIL for provision of synchrotron radiation facilities and we would like to thank Aurélien Thureau for assistance in using beamline SWING and help with complementary SAXS measurements. 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. We finally thank the reviewers for their thorough and critical analysis of the manuscript and their constructive comments.

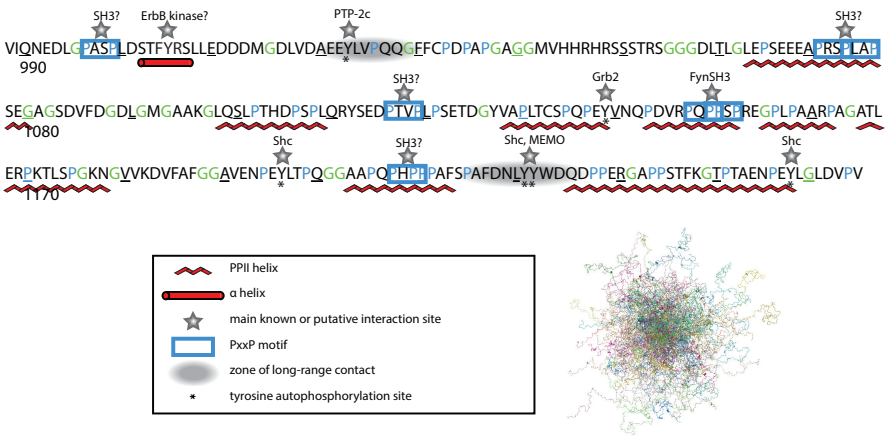

Figure 3: Model of CtErbB2 structural features with some highlighted functional sites. *Along the sequence:* The transient secondary structures revealed by SSP scores, residual dipolar couplings and scalar couplings are indicated, as well as the main fragments exhibiting transient long-range contacts observed in the PRE experiments. In addition to the main known interaction sites and corresponding partners (Shc and Grb2 (25, 26), PTP-2c (26) and MEMO(27) for interaction with phosphotyrosines, FynSH3 (87) 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 the first 20 N-terminal residues of the construct.

REFERENCES

1. Lemmon, M. A., and J. Schlessinger, 2010. Cell Signaling by Receptor Tyrosine Kinases. Cell 141:1117–1134. https://linkinghub.elsevier.com/retrieve/pii/S0092867410006653.

2. Roskoski, R., 2004. The ErbB/HER receptor protein-tyrosine kinases and cancer. Biochemical and Biophysical Research Communications 319:1–11. https://linkinghub.elsevier.com/retrieve/pii/S0006291X04008757.

3. Lemmon, M. A., J. Schlessinger, and K. M. Ferguson, 2014. The EGFR family: not so prototypical receptor tyrosine kinases. Cold Spring Harbor Perspectives in Biology 6:a020768.

4. 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. The Journal of Biological Chemistry 266:13828–13833.

5. Yarden, Y., and M. X. Sliwkowski, 2001. Untangling the ErbB signalling network. Nature Reviews Molecular Cell Biology 2:127–137. http://www.nature.com/articles/35052073.

6. 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. https://linkinghub.elsevier.com/retrieve/pii/S1097276503000480.

7. Slamon, D., G. Clark, S. Wong, W. Levin, A. Ullrich, and W. McGuire, 1987. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235:177–182. https://www.sciencemag.org/lookup/doi/10.1126/science.3798106.

8. 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. http://link.springer.com/10.1007/s10911-006-9009-1, tex.ids= badache2006.

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

10. 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.

11. 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.

12. 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.

13. 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 p185 Protein of Transformed Cells by the Antibody chA21*. Journal of Biological Chemistry 286:31676–31683. https://linkinghub.elsevier.com/retrieve/pii/S0021925820722751.

14. 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.

15. 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 (London, England: 1993) 25:878–889.e5.

16. 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.

17. 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.

18. 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. ISBN: 1878-4186 (Electronic)\r0969-2126 (Linking).

19. 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. ISBN: 1091-6490 (Electronic)\n0027-8424 (Linking).

20. 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. ISBN: 0021-9258 (Print)\r0021-9258 (Linking).

21. 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. Tex.ids= Mineev2010 ISBN: 1089-8638 (Electronic)\r0022-2836 (Linking) publisher: Elsevier Ltd.

22. 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. Publisher: Elsevier Ltd ISBN: 1089-8638.

23. 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. https://linkinghub.elsevier.com/retrieve/pii/S0021925820511332, tex.ids= Aertgeerts2011 ISBN: 1083-351X.

24. 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. Tex.ids= Ishikawa2011 ISBN: 9789513944926.

25. 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. ISBN: 0270-7306 (Print)\r0270-7306 (Linking).

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

27. 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.

28. 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. ISBN: 0021-9258 (Print)\r0021-9258 (Linking).

29. Okamoto, K., and Y. Sako, 2019. Single-Molecule Förster Resonance Energy Transfer Measurement Reveals the Dynamic Partially Ordered Structure of the Epidermal Growth Factor Receptor C-Tail Domain. The Journal of Physical Chemistry B 123:571–581. https://doi.org/10.1021/acs.jpcb.8b10066, publisher: American Chemical Society tex.ids= okamoto2018.

30. Regmi, R., S. Srinivasan, A. P. Latham, V. Kukshal, W. Cui, B. Zhang, R. Bose, and G. S. Schlau-Cohen, 2020. Phosphorylation-Dependent Conformations of the Disordered Carboxyl-Terminus Domain in the Epidermal Growth Factor Receptor. The Journal of Physical Chemistry Letters 10037–10044. https://pubs.acs.org/doi/10.1021/acs.jpclett.0c02327.

31. 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. ISBN: 0961-8368 (Print)\r0961-8368 (Linking).

32. 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: A Publication of the Protein Society 15:1142–1152. Tex.ids= Lee2006 ISBN: 0961-8368 (Print)\r0961-8368 (Linking).

33. Gill, K., J. L. Macdonald-Obermann, and L. J. Pike, 2017. Epidermal growth factor receptors containing a single tyrosine in their C-terminal tail bind different effector molecules and are signaling-competent. Journal of Biological Chemistry 292:20744–20755. https://linkinghub.elsevier.com/retrieve/pii/S0021925820328064, tex.ids= Gill2017.

34. 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.

35. 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. The Journal of Biological Chemistry 292:597–610. Tex.ids= Keppel2017.

36. Wright, P. E., and H. J. Dyson, 2015. Intrinsically disordered proteins in cellular signalling and regulation. Nature reviews. Molecular cell biology 16:18–29. ArXiv: 15334406 ISBN: 0022219412.

37. 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 of the United States of America 97:8868–8873. Tex.ids= Shoemaker2000 ISBN: 0027-8424\r1091-6490.

38. 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. Tex.ids= Huang2009 ISBN: 1089-8638 (Electronic)\r0022-2836 (Linking).

39. 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: PCCP 14:10466–10476. Tex.ids= Zhou2012 arXiv: NIHMS150003 ISBN: 1463-9076.

40. 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. https://linkinghub.elsevier.com/retrieve/pii/S1359644618300448, tex.ids= Ruan2018 publisher: Elsevier Ltd.

41. 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. Tex.ids= Sibille2012 ISBN: 0300-5127.

42. 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. ISBN: 0887-3585.

43. 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. ISBN: 0002-7863.

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

45. DeJean de la Batie, R., F. Laupretre, and L. Monnerie, 1988. Carbon-13 NMR investigation of local dynamics in bulk polymers at temperatures well above the glass transition temperature. 1. Poly(vinyl methyl ether). Macromolecules 21:2045–2052. https://pubs.acs.org/doi/abs/10.1021/ma00185a028, tex.ids= Dejean1988a.

46. DeJean de la Batie, R., F. Laupretre, and L. Monnerie, 1988. Carbon-13 NMR investigation of local dynamics in bulk polymers at temperatures well above the glass transition temperature. 2. Poly(propylene oxide) and linear and cross-linked poly(ethylene oxides). Macromolecules 21:2052–2058. https://pubs.acs.org/doi/abs/10.1021/ma00185a029, tex.ids= Dejean1988b.

47. Bahar, I., B. Erman, and L. Monnerie, 1994. Effect of molecular structure on local chain dynamics: Analytical approaches and computational methods. In A. Abe, H. Benoit, H.-J. Cantow, P. Corradini, K. Dušek, S. Edwards, H. Fujita, G. Glöckner, H. Höcker, H.-H. Hörhold, H.-H. Kausch, J. P. Kennedy, J. L. Koenig, A. Ledwith, J. E. McGrath, L. Monnerie, S. Okamura, C. G. Overberger, H. Ringsdorf, T. Saegusa, J. C. Salamone, J. L. Schrag, G. Wegner, L. Monnerie, and U. W. Suter, editors, Atomistic Modeling of Physical Properties, Springer Berlin Heidelberg, Berlin, Heidelberg, volume 116, 145–206. http://link.springer.com/10.1007/BFb0080199, tex.ids= Bahar1994 seriesTitle: Advances in Polymer Science.

48. 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.

49. 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 7:269. Tex.ids= Avdoshenko2017.

50. 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.

51. 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. ISBN: 0925-2738 (Print) 0925-2738 (Linking).

52. 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. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts172, tex.ids= Ozenne2012 ISBN: 1367-4803.

53. 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. ISBN: 0002-7863.

54. 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. ISBN: 1090-7807 (Print)\r1090-7807 (Linking).

55. 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.

56. Sharp, P., and V. A. Bloomfield, 1968. Light scattering from wormlike chains with excluded volume effects. Biopolymers 6:1201–1211. ISBN: 0006-3525 (Print)\r0006-3525 (Linking).

57. Pedersen, J. S., and P. Schurtenberger, 1996. Scattering functions of semiflexible polymers with and without excluded volume effects. Macromolecules 29:7602–7612. ISBN: 0024-9297.

58. 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. Publisher: Nature Publishing Group ISBN: doi:10.1038/srep14223.

59. 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. Tex.ids= Receveur-Brechot2012 ISBN: 1875-5550 (Electronic)\r1389-2037 (Linking).

60. Lopes, J. L. S., A. J. Miles, L. Whitmore, and B. A. Wallace, 2014. Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: applications in secondary structure analyses. Protein Science: A Publication of the Protein Society 23:1765–1772. Tex.ids= Lopes2014 ISBN: 1469-896X (Electronic) 0961-8368 (Linking).

61. 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. Tex.ids= Bernado2009 ISBN: 0006-3495 publisher: Biophysical Society.

62. 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. Tex.ids= Wilkins1999 ISBN: 0006-2960.

63. 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: A Publication of the Protein Society 15:2795–2804. Tex.ids= Marsh2006 ISBN: 0961-8368 (Print)\r0961-8368 (Linking).

64. Bernadó, 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 of the United States of America 102:17002–17007. Tex.ids= Bernado ISBN: 0027-8424 (Print)\r0027-8424 (Linking).

65. Mohana-Borges, R., N. K. Goto, G. J. Kroon, H. Dyson, and P. E. Wright, 2004. Structural Characterization of Unfolded States of Apomyoglobin using Residual Dipolar Couplings. Journal of Molecular Biology 340:1131–1142. https://linkinghub.elsevier.com/retrieve/pii/S0022283604005881, tex.ids= Mohana2004 ISBN: 0022-2836 (Print)\n0022-2836 (Linking).

66. 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. Tex.ids= Fuxreiter2004 ISBN: 0022-2836 (Print).

67. 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. Tex.ids= Ferreon ISBN: 0961-8368 (Print)\r0961-8368 (Linking).

68. Lam, S. L., and V. L. Hsu, 2003. NMR identification of left-handed polyproline type II helices. Biopolymers 69:270–281. http://doi.wiley.com/10.1002/bip.10354, tex.ids= Lam2003.

69. Smith, L. J., K. A. Bolin, H. Schwalbe, M. W. MacArthur, J. M. Thornton, and C. M. Dobson, 1996. Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. Journal of Molecular Biology 255:494–506. Tex.ids= Smith1996 ISBN: 0022-2836.

70. 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. ISBN: 7654961189.

71. Serrano, L., and A. R. Fersht, 1989. Capping and α-helix stability. Nature 342:296–299. http://www.nature.com/articles/342296a0, tex.ids= Serrano1989.

72. Jensen, M. R., G. Communie, E. A. 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 of the United States of America 108:9839–9844. Tex.ids= Jensen2011 ISBN: 1091-6490 (Electronic)\r0027-8424 (Linking).

73. 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 (New York, N.Y.) 295:1719–1722. Tex.ids= Klein-Seetharaman2002 ISBN: 0036-8075.

74. Khan, S., C. Charlier, R. Augustyniak, N. Salvi, V. Déjean, 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. https://linkinghub.elsevier.com/retrieve/pii/S0006349515007316, tex.ids= Khan2015.

75. Gast, K., H. Damaschun, K. Eckert, K. Schulze-Forster, H. R. Maurer, M. Müller-Frohne, D. Zirwer, J. Czarnecki, and G. Damaschun, 1995. Prothymosin alpha: a biologically active protein with random coil conformation. Biochemistry 34:13211–13218. Tex.ids= Gast1995 ISBN: 1321113218.

76. 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. Tex.ids= Schwalbe1997 ISBN: 0006-2960 (Print)\r0006-2960 (Linking).

77. 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. Tex.ids= Kovacs2015a.

78. 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. Tex.ids= Zhang2006 arXiv: NIHMS150003 ISBN: 0092-8674 (Print)\n0092-8674 (Linking).

79. 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. Tex.ids= Pines2010 ISBN: 1476-5594 (Electronic)\r0950-9232 (Linking) publisher: Nature Publishing Group.

80. Sorkin, A., M. Mazzotti, T. Sorkina, L. Scotto, and L. Beguinot, 1996. Epidermal Growth Factor Receptor Interaction with Clathrin Adaptors Is Mediated by the Tyr974-containing Internalization Motif. Journal of Biological Chemistry 271:13377–13384. https://linkinghub.elsevier.com/retrieve/pii/S0021925818413336.

81. 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.

82. 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.

83. 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. ISBN: 1932-6203 (Electronic)\n1932-6203 (Linking).

84. Pawson, T., 1992. SH2 and SH3 domains. Current Opinion in Structural Biology 2:432–437. ISBN: 0960-9822 (Print).

85. 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. ISBN: 0014-5793 (Print)\r0014-5793 (Linking).

86. Saksela, K., and P. Permi, 2012. SH3 domain ligand binding: What's the consensus and where's the specificity? FEBS Letters 586:2609–2614. ISBN: 0014-5793.

87. 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.

88. 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.

89. 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. ISBN: 0261-4189 (Print).

90. 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. ISBN: 0950-9232.

91. Birrane, G., J. Chung, and J. A. A. Ladias, 2003. Novel Mode of Ligand Recognition by the Erbin PDZ Domain *. Journal of Biological Chemistry 278:1399–1402. https://www.jbc.org/article/S0021-9258(19)31049-X/abstract, publisher: Elsevier.

92. Khurshid, R., M. Saleem, Gul-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.

93. Mateos, B., C. Conrad-Billroth, M. Schiavina, A. Beier, G. Kontaxis, R. Konrat, I. C. Felli, and R. Pierattelli, 2020. The Ambivalent Role of Proline Residues in an Intrinsically Disordered Protein: From Disorder Promoters to Compaction Facilitators. Journal of Molecular Biology 432:3093–3111. https://linkinghub.elsevier.com/retrieve/pii/S0022283619306801, tex.ids= Mateos2020.

94. Rice, P., L. Longden, and A. Bleasby, 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics 16:276–277. ArXiv: NIHMS150003 ISBN: 1359-0294.

95. Elam, W. A., T. P. Schrank, A. J. Campagnolo, and V. J. Hilser, 2013. Temperature and urea have opposing impacts on polyproline II conformational Bias. Biochemistry 52:949–958.

SUPPLEMENTARY MATERIAL

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