bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cadherin-11 dimerization multi-site kinetics: combined partial unfolding and strand-swapping

Hans Koss[1], Barry Honig[1,2,3,4], Lawrence Shapiro[1,2], Arthur G Palmer III[1*]

[1]Department of Biochemistry and Molecular Biophysics
Columbia University Irving Medical Center
701 West 168th Street
New York, NY 10032, USA

[2]Zuckerman Brain Mind Behaviour Institute
Columbia University
3227 Broadway
New York, NY 10027, USA.

[3]Department of Systems Biology
Columbia University Irving Medical Center
1130 St. Nicholas Avenue
New York, NY, 10032, USA.

[4]Department of Medicine
630 West 168th Street
Columbia University Irving Medical Center
New York, NY, 10032, USA.

Address correspondence to A.G.P.:
Telephone: 212 305-8675
Email: agp6@columbia.edu

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Cadherin extracellular domain 1 (EC1) mediates dimerization of type II cadherins between opposing cell surfaces to facilitate cell-cell adhesion. EC1 forms domain-swapped dimers, with residues Trp2 and Trp4 crucial to anchor the EC1 A-strand in the strand-swapped conformation. In the present work, the specific roles of Trp2 and Trp4 in the dimerization mechanism of Cadherin-11 have been elucidated using NMR spectroscopy of wild-type and designed mutant EC1 domains. The results show that the dominant monomeric state, with the A strand and Trp side chains packed intramolecularly, is in equilibrium with a sparsely populated (1.6%) partially strand-exposed state, in which the Trp2 side chain packing is disrupted and with a sparsely populated (1.6%) fully strand-exposed state, in which the A strand, Trp2 and Trp4 side chain packing are fully disrupted. The exchange kinetics between the major state and the partially strand-exposed state are slow-intermediate (k_{ex} = 700 – 734 s^{-1}) and the exchange kinetics between the major state and the fully strand-exposed state are intermediate-fast (k_{ex} = 3470 - 3591 $s^{-1,}$) on the NMR chemical shift time scale. These three conformations, varying in the degree of A-strand exposure, also are coupled to additional conformational states on very fast and very slow timescales. The very fast exchange process arises from interconversion between ordered and random coil conformations of the BC loop in proximity to the Trp2 binding pocket, with relative populations that depend on the extent of A-strand exposure and dimerization status. The very slow exchange processes link the folded partially and fully strand-exposed conformations with partially unfolded conformational states, which have been recognized as crucial intermediates for domain-swapping in proteins other than cadherins. This framework reveals the dimerization mechanism of type II Cadherins as coupled folding and strand-swapping.

Introduction

Classical cadherins, which include the type I and type II families in vertebrates, are calcium-dependent cell-cell adhesion proteins [1]. Type II cadherin ectodomains contain five “extracellular cadherin” (EC) domains. The membrane-distal cadherin EC1 domains from cadherin molecules on the surfaces of opposing cells form dimers in the process of cell-cell adhesion. The greatly varying dimerization propensities of various cadherin homo- and heterodimers is thought to play an important role in tissue patterning [2]. EC1 is the primary determinant of type II cadherin dimerization specificity; the high level of similarity of dimeric interfaces in type II Cadherins raises the question of how this specificity is achieved [3]. While isolated EC1 constructs can form stable dimers, and all intermolecular contacts are located between EC1 domains, binding of Ca^{2+} rigidifies the connections between successive EC domains

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and dimerization of EC1-EC2 protein constructs is facilitated by formation an "X-dimer" kinetic intermediate can [4, 5].

Type II cadherin EC1 domains form strand-swapped dimers, with the N-terminal A-strand undocking from an intramolecular site in the monomer to bind an identical intermolecular site in the partner protomer [3, 6]. The side chains of residues Trp2 and Trp4 near the N-terminus are packed intramolecularly in the monomer conformation and intermolecularly to anchor the A strand in the domain-swapped conformation [3]. Prior work has suggested that Cadherin 8, a type II Cadherin, can access a strand-exposed conformation in the monomer state to facilitate dimerization [7]. However, the existence of a strand-exposed intermediate in Cadherin dimerization is still elusive and the precise role of Trp4 in type II Cadherins is unknown [7, 8].

Partially or fully unfolded intermediates have been described for many domain-swapping dimerization processes, compatible with the concept that domain swapping is intermolecular protein folding [9, 10]. Conformational heterogeneity, including partially folded species, has been detected in a β-strand swapping GB1 mutant [11]. Molecular dynamics simulations and experimental studies of RNase A suggest that C-terminal domain-swapping proceeds via substantially unfolded intermediate states [12, 13]. Domain swapping can require full unfolding, as exemplified for Cyanovirin-N protein [14]. Unfolding can promote domain swapping in a variety of ways. (1) High concentrations may favor non-thermal 'melting' of the protein in which a protein is acting as a solvent [15, 16]. Large hydrophobic surfaces could promote the formation of encounter complexes, as they occur in induced-fit mechanisms [17]. Accordingly, the presence of monomeric and dimeric intermediates of M^{pro}-C of SARS-CoV main protease in which the C-terminal A5-helix is unfolded may provide a hydrophobic environment to enable a1-helix strand-swapping [18, 19]. (2) Simulations of the strand-swapping SH3 domain of epidermal growth factor receptor pathway substrate 8 (Eps8) show that unfolded intermediate states increase competition between intra- and inter-chain interactions by virtue of their encounter complex interfaces [20]. (3) Unfolded states can serve as free energy traps in the domain-swapping process while avoiding an overly rugged energy landscape during dimerization [20].

The present work characterizes the conformational states of the monomeric form of the Type II cadherin-11 EC1 (Cad11-EC1) domain[3, 21, 22]. Extensive NMR spectroscopic measurements, including chemical shift perturbation and relaxation dispersion, were conducted on wild-type (WT) and designed mutants of Cad11-EC1, including dependence on temperature and pressure. The results describe an ensemble of folded, locally disordered, A-strand exposed, and partially unfolded conformational states

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of Cad11-EC1 in solution. This landscape of conformations interconverting at different time scales reveals dimerization of Cadherin Type II domains as coupled folding and A-strand swapping and emphasizes the importance of conformational plasticity of the Trp2 binding pocket and the destabilizing influence of Trp4 in the mechanism.

Results

The key NMR probes of the conformational dynamics of the Cadherin-11-EC1 (WT-C11) domain are Trp2 and Trp4 in the A-strand of the domain, together with the flanking BC (residues 29-34), DE (residues 52-55), and FG (residues 80-84) loops, as shown in ***Fig. 1a***.

Monomer and dimer NMR spectra

The ^{1}H-^{15}N correlation spectrum of Cad11-EC1 (***Fig. 2a***) displays substantial line broadening and, in many cases, more than one resonance peak per residue. By comparing spectra at different concentrations, we were able to assign resonance peaks to a main monomer state, a main dimer state, or an overlap of each. Assignments of resonances to monomer or dimer states were confirmed by comparing experimentally obtained relaxation rate constant ratios with those calculated from the crystal structure (***Suppl. Fig. 1***). A total of 42 residues (46 % excluding prolines) were assigned for the main monomer state of WT Cad11-EC1. The two-domain Cadherin-11-EC1-EC2 protein construct is difficult to access under conditions suitable for NMR spectroscopy. However, the overlap between ^{1}H-^{15}N-HSQC spectra for Cad11-EC1 and Cadherin-11-EC1-EC2 indicates that the properties of Cad11-EC1 are similar in the two-domain context (***Suppl. Fig. 2***).

Trp2 and Trp4 mutants reveal very fast dynamics between ordered and random coil states of the BC loop.

We introduced W2A and/or W4A mutants to test the role of individual Trp (W) residues. New resonance peaks were observed in NMR spectra of the mutant domains. A subset of the new resonance peaks corresponds to the random coil (RC) states: the correlation between predicted random coil ^{15}N-shifts and assigned ^{15}N resonances is 96% (***Suppl. Fig. 3***). The ^{1}H,^{15}N-HSQC peaks for BC loop residues Ser30-Gly33 in wild-type and mutant domains can be placed on a line connecting the peaks for WT-dimer and RC monomer (***Fig. 1b, Suppl. Fig. 4***). This correlation indicates that a structurally ***ordered*** state kinetically exchanges with a disordered ***random coil*** (RC) state. The resonance peaks for dimer states are shifted in the opposite direction from the RC states. Thus, the fraction of RC state (equal to the degree of ***local unfolding***) in the BC loop generally follows the order: WT(dimer) – WT - W4A - W2A- W2AW4A -

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RC, with WT-dimer being predominantly ordered (states without a designation as dimer are monomeric). The RC state population varies for each mutant (***Fig. 1c***), suggesting that the fast exchange equilibria are partially independent. The RC population of the W2AW4A mutant obtained from $^{13}C_{\alpha}$ resonance peak positions for residues Ser30, Asp32 and Gly33 match the results from the analysis of ^{1}H-^{15}N spectra (***Suppl. Fig. 4***). Resonance peak broadening due to this fast exchange process could not be suppressed in high power relaxation dispersion experiments (*vide infra*); therefore, the exchange process is very fast on the NMR chemical shift time scale (< 10 - 100 μs). The peak patterns observed for residues in the DE and FG loop, particularly Asp81 and Asn83, in WT and mutant C11 could suggest a very fast exchange kinetic process a well; however, the linear correlation of the chemical shifts for the wild-type and mutant proteins does not extend to the RC peak positions. Alternatively, the peak patterns in the DE and FG loops could be a consequence of local chemical shift perturbations by the structural changes in the neighboring BC loop.

Population and structure of partially unfolded states depends on Trp status

The Cad11-EC1 monomer structures associated with RC resonance frequencies are ***partially unfolded states.*** The resonance peaks for folded states are resolved from the RC peaks and hence the folded and partially folded states exchange slowly on the NMR chemical shift time scale. The overall intensities of RC peaks for the W2A mutant are 4.3 times smaller than for the W2AW4A mutant when scaled to the average folded state intensity, showing increased population the partially folded state for the W2AW4A mutant.

RC peak intensities for the mutant proteins are greatly influenced by the populations of these states and by the effects of the local structural context on effective rotational correlation times. Therefore, we normalized RC peak intensities by the average RC intensity of all assigned residues up until and including DE loop (residues 1-55). This normalization yielded identical residue-wise intensity profiles for W2A-C11 and W2AW4A-C11 (***Fig. 1d and Suppl. Fig. 5a***). This agreement indicates that large regions of partially unfolded states of W2A and W2AW4A mutants are structurally identical, pointing to a common unfolding pathway. However, resonances for residue from Val78 to Ile93, including the FG loop, generally display greater normalized RC peak intensities in the W2AW4A mutant. Correspondingly, the region including FG loop appears particularly destabilized in the W4A mutant (***Suppl. Fig. 5b***). These two results suggest that the W4A mutation promotes unfolding in the region of the FG loop. In summary, the fast kinetic equilibrium between RC and ordered states is highly localized between BC and FG loops. The

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

scheme in ***Fig. 1e*** encapsulates both exchange processes, also suggesting regional structural similarities in partially unfolded states.

High pressure NMR experiments

The ^{1}H,^{15}N-HSQC spectra obtained from static high-pressure titrations from 1 bar to 2.07 kbar show dramatic changes in numbers, intensities, and frequencies of resonance peaks (***Figure 2, Suppl. Fig. 6***). Peak intensity and frequency changes are illustrated in Figures 3 and 4. An overview of the various pressure-dependent species is given in ***Fig. 3f***. Pressure-dependent intensities of RC and folded peaks reveal unfolding under pressure. Increases in folded monomer peak intensities at low pressures arise in part from pressure-dependent dimer-monomer transitions; however, local packing effects or chemical exchange processes further modulate site-specific peak intensities.

Multi-step pressure unfolding reveals intermediate- and high-pressure partially unfolded states

At 1.72 kbar, the ^{1}H,^{15}N- HSQC spectrum is consistent with a mostly unfolded state of WT Cad11-EC1 (***Figure 2, Suppl. Fig. 6***), in which the resonance frequencies for all assigned residues agree with RC values. Correspondingly, ~20 new pressure-dependent resonance peaks match the assignments of the RC peaks for the W2AW4A mutant.

Pressure-dependent resonance peak intensities and frequencies were fit with a generic spline function that depends on two linear features and a critical pressure at which the linear slopes change (*Online Methods*). The behavior of the intensities and resonance positions of RC peaks suggests that unfolding from 0 kbar to 1.72 kbar occurs in at least two distinct stages, from 0-0.97 kbar and 0.97-1.72 kbar, thus defining *intermediate-pressure (IP-UNF)* and *high-pressure (HP-UNF) partially unfolded* states (***Figs. 2, 3 a-c***, ***Suppl. Fig. 7-8***). Resonance peaks for the *IP-UNF* state shift uniformly towards the resonance positions for the *HP-UNF* state at pressures > 0.97 kbar, commensurate with fast exchange between these states. The linear correlation between the spectra for the wild-type domain at pressures < 0.97 kbar and the W2AW4A mutant at 1 bar suggest that the IP-UNF structure is similar to the partially unfolded W2AW4A state.

Pressure-dependence of unfolded peak intensities and of folded peak chemical shifts indicates locally varying stability

Pressure-dependent shifts in the ^{1}H and ^{15}N resonance frequencies for peaks associated with folded conformations arise from shifts in the populations of alternative states that are in fast exchange on the

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

NMR chemical shift timescale (***Figs. 3e, Suppl. Figs. 7-9***). Resonances for folded conformations of residues 31-34 in the BC loop peaks shift towards the HP-UNF peak positions as pressure increases, suggesting that the HP-UNF state is locally accessible for BC loop residues even at pressures < 0.97 kbar. In contrast, the peak positions for residue Ser54 in the DE loop and residues Asp81 and Asn83 in the FG loop shift non-linearly with increasing pressure and only towards the HP-UNF peak positions only at high pressures ($\gtrapprox$ 0.97 kbar, for Ser54 at 298K; ***Fig. 3e, Suppl. Figs. 7-9***), notwithstanding an intensity decrease at $\gtrapprox$ 0.69 bar due to a diminishing folded protein population ***(Fig. 3d)***. This result also suggests that the DE and FG loops are in fast-exchange with two ordered conformations, consistent with the mutational results, and furthermore is more pressure-stable than the BC loop.

Relaxation dispersion experiments for multi-site kinetics

Carr-Purcell-Meiboom-Gill (CPMG), CEST, Hahn-echo, and high power spin-lock ^{15}N NMR experiments for WT Cad11-EC1 were used to characterize multi-site kinetics (***Fig. 4***) over a wide range of µs-ms timescales, while reducing ambiguity in the subsequent multi-site fitting time scales (Online methods) [23].

Three-site model for strand exposure kinetics

Global three-site fits were performed for 12 residues ***(Fig. 4, Suppl. Fig. 10, Suppl. Fig. 11)***. We find slow-intermediate exchange (k_{ex} = 700 – 734 s^{-1}) between a major (A-strand-bound) monomeric state (population: 96.8 %) and a partially strand-exposed sparsely populated state (1.62-1.65 %) and intermediate-fast (k_{ex} = 3470 - 3591 s^{-1}) exchange between the major state and a fully strand-exposed sparsely populated state (1.60-1.63 %) ***(Fig. 6)***. The identified chemical shifts of the partially strand-exposed state map to those of the folded conformation of the W2A mutant, suggesting that this mutant is a model of the partially-strand exposed state of the WT domain. The shifts of the fully strand-exposed states mostly map to those of the folded conformation of the W2AW4A mutant with a slightly higher RC share, suggesting that the W2AW4A mutant is only mostly strand-exposed. The model is supported by qualitative analysis of ^{1}H-CPMG relaxation data recorded for [U-^{2}H, U-^{15}N]-labeled samples (***Suppl. Fig. 12***). The data do not resolve whether direct exchange occurs between the partially and fully strand-exposed states (yielding a triangular kinetic scheme).

Concentration dependence of fast-intermediate exchange between strand-bound and fully strand-exposed

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Many chemical shifts in ^{1}H-^{15}N correlation spectra for Cad11-EC1 display a strong B_0 field dependence (***Suppl. Fig. 13a***), which itself is a signature of chemical exchange. The B_0-induced chemical shift perturbations between 500 and 900 MHz ($\Delta\delta_{B0}$) correlate linearly between low (160 mM, $\Delta\delta_{B0,low}$) and high (640 mM, $\Delta\delta_{B0,high}$) concentration samples, ruling out most artifacts, systemic and sample-related errors contributing to the shift perturbations (***Suppl. Figs. 13c, 13d***). The average B_0-dependent shift perturbation is 10% greater at high concentration, suggesting that the underlying chemical exchange process has a degree of concentration dependence. The B_0-dependent ^{15}N chemical shift differences at high concentration also are correlated (R = 0.72) with the ^{15}N chemical shift differences between the major conformation and fully-strand exposed conformation detected in relaxation dispersion experiments (***Suppl. Fig. 13b***).

Temperature dependence reveals features of partially unfolded states

We recorded relaxation dispersion at 285 K and 298 K. The slow-intermediate process between WT and the partially strand-exposed (W2A-like) state is temperature independent. The slow-intermediate process was mapped to the structure by directly comparing the peak intensities in ^{1}H,^{15}N-HSQC spectra recorded at different fields and temperatures (***Fig. 5***); no chemical shifts or data obtained from relaxation measurements were needed to perform this procedure. We find that the slow-intermediate process indeed maps to the W2A binding pocket. The line-broadening arising from exchange between WT and the fully strand-exposed state (W2W4-like) decreases with increasing temperature, as expected for an intermediate-fast time scale kinetic process. As already noted, the fully strand-exposed state is closely identified with the folded W2AW4A-like mutant and the WT IP-UNF states.

Discussion and Conclusion

As a type II Cadherin, Cadherin-11 carries Trp2 and Trp4 as two key residues involved in strand swapping. To investigate the roles of these Trp residues in the mechanism of strand swapping, we quantitatively characterized fast-intermediate timescale chemical exchange of WT-Cad11 between the main strand-bound and a sparsely populated fully strand-exposed (W2AW4A mutant-like) state, and slow-intermediate timescale chemical exchange between the main strand-bound and a sparsely populated partially strand-exposed (W2A mutant-like) state (***Fig. 6***). The fully stand-exposed state also has both Trp2 and Trp4 unpacked from the intramolecular binding pocket, while the partially exposed strand has only Trp2 unpacked. Additional very fast exchange occurs between ordered and disordered (random-coil like) conformations of the BC loop, while the neighboring DE and FG loops are either in fast exchange between two alternative ordered conformations or structurally influenced by the unfolding process of

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the BC loop. The populations of the RC states of loops increase in the order dimer < monomer strand-bound < WT-partially strand-exposed < fully strand-exposed. Because exchange between strand-bound and partially strand-exposed states is slower than between strand-bound and fully strand-exposed states, we conclude that Trp2 exposure is not a required precursor to Trp2 and Trp4 exposure. Rather, Trp2 and Trp4 detach from the core either simultaneously or via a Trp4-exposed state.

Mutation studies suggest that unbinding Trp2 and/or Trp4 increases RC populations in BC loop. However, the strand-bound dimer state leads to a larger ordered population than the strand-bound monomer state and hence reflects stabilization of the dimer state compared to the monomer state. This implies that the maximum ordered population of the dimer is not merely a result of stabilization by the strand-bound state. This stabilization of the ordered state of BC loop in the dimer, but not in the monomer, might be a factor, in addition to conformational strain differences in the A strand[24], to rationalize the preference for strand-swapping over self-insertion of the strand.

Very slow chemical exchange of folded states with partially unfolded states also depends on strand conformation and packing of Trp2 and Trp4. The effects of Trp2 and Trp4 exposure distinctly target different sections of the EC1 domain. Thus, FG loop residues are destabilized more by exposure of Trp4 than by exposure of Trp2. We suggest a direct kinetic connection between the fully strand-exposed state and a partially unfolded state that is structurally similar to the intermediate pressure state. Pressure-dependent fast exchange between folded and unfolded states support the suggestion that at least the partially unfolded state is accessible at ambient pressure. Trp4 exposure in isolation only marginally increases the very fast exchanging random coil population in the BC loop; the combined exposure with Trp2, in contrast, leads to a dramatic increase of both a very slow exchanging partially unfolded state and a random coil state. We conclude that any local destabilization of the BC loop region by Trp4 occurs probably via the very slow exchanging partially unfolded state. A higher population of the partially unfolded state may also promote formation or stabilization of the fully strand-exposed state. These newly discovered partially unfolded states and local unfolding, in conjunction with strand-swapping, provide the kinetic framework to stabilize dimer-strand-binding over self-insertion of the strand, as described in other domain-swapping proteins [11, 12, 14, 16-19][.

Modulation of Cadherin-11 monomer stability could, in principle, be fine-tuned by a dimerization partner or by intermolecular contacts in the X-dimer. The contact surface in type II Cadherins is particularly large and hydrophobic. Similar to other strand-swapping proteins, partially unfolded regions

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

or hydrophobic surfaces could support encounter complex formation, or be viewed as "protein solvent" to catalyze domain swapping, conceptualized as an intermolecular folding process.

According to the Le Chatelier principle, for a single slow dimerization process, the relative populations of monomer states participating in exchange processes faster than dimerization are not concentration dependent. Therefore, the concentration dependence of the observed chemical exchange processes implies not only that the partially strand exposed state is on pathway for Cadherin-11 dimerization, but also that more than one dimerization pathway exists (***Suppl. Fig. 14***) [7]. At high concentrations, as encountered in the 2D-network of Cadherins organizing to form adherens junctions, encounter complexes involving partially folded proteins acting as "protein solvent" might be advantageous.

The present work has revealed a complex hierarchy of sparsely populated conformational states of the Cadherin-11 EC1 domain. These states, ranging from localized alternative conformations of loops surrounding the Trp binding pockets to partially or fully A-strand exposed states (with exposure of Trp2, Trp4 or both) to more extensively partially unfolded conformations, are a central aspect of Cadherin-11, and probably other type II Cadherin, function. Trp4 is present in type II, but not type I cadherins. We found Trp4 exposure in the fully strand-exposed state to give a larger partially unfolded state population and local destabilization at least of the FG loop region. In addition to the common advantages of unfolded intermediates in domain-swapping proteins, unfolded intermediate states might offer an evolutionary advantage for fine-tuning differential Cadherin dimerization specificity.

Acknowledgements

We acknowledge Shibani Bhattacharya (New York Structural Biology Center, NYSBC), Mike Goger (NYSBC), Julia Brasch (Columbia University) and Anna Kacyznska (Columbia University) for helpful discussions. This work was supported by National Institutes of Health grant R35 GM130398 (A.G.P.), by National Institutes of Health grant R01 MH11481708 (L.S.) and by the US National Science Foundation grant MCB-1412472 (to B.H.). Some of the NMR experiments were conducted at the Center on Macromolecular Dynamics by NMR Spectroscopy located at NYSBC, supported by a grant from the NIH National Institute of General Medical Sciences (GM118302). B.H., L.S., and A.G.P. are members NYSBC. The data collection at NYSBC was made possible by a grant from ORIP/NIH facility improvement grant CO6RR015495. Data collected using the 800 MHz Avance III NMR spectrometer is supported by NIH grant S10OD016432. The 700 MHz NMR spectrometer was purchased with funds from NIH grant S10OD018509. The 900 MHz NMR spectrometers were purchased with funds from NIH grant P41GM066354 and the New York State Assembly.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Author Contributions

B.H., L.S., and A.G.P. conceptualized the research program. H.K. and A.G.P. designed NMR experiments. H.K. conducted research, wrote the data analysis scripts, and analyzed the data. All authors contributed to interpretation of the results and writing of the paper.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Online Methods

1 Protein expression and purification

Plasmids for overexpression of WT Cad11-EC1 (residues 1-98) and Cad11-EC1EC2 (Residues 1-207) were constructed using a pSMT3 vector [3]. The constructs included an N-terminal 6-HIS-SUMO-tag . Plasmids for W2A, W4A and W2AW4A mutants were obtained commercially (Genewiz). WT plasmids were generously provided by Julia Brasch and Anna Kaczynska (L.S. lab, Columbia University).

The following procedure was used to produce ^{15}N or ^{13}C, ^{15}N-labelled protein. The pSMT vector harboring the construct was transformed into E. coli strain C41 (DE3) on 2xYT/Kanamycin plates. Eight colonies were picked and incubated to 37 °C overnight in 5 ml 2xYT/Kanamycin. Glycerol stocks were prepared from the overnight cultures. Cultures were inoculated from glycerol stocks and grown until plateauing at high density (OD ~ 4) in 1-3.5 l 2xYT/Kana (0.5 l / flask) at 37 °C. Following a protocol related to Murray et al. [25],cells were pelleted and carefully resuspended in a 2l flask with 500 ml minimal medium (90 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.5 mM NaCl, 18.7 mM ^{15}N-NH_4Cl, 28 mM D-Glucose (^{13}C-Glucose: 11 mM), 8.7 mM Na_2SO_3, 0.5 x Trace Metal Mix, 50 µg/ml Kanamycin, 20 mg/l Thiamin, 20 mg/l Biotin, 2 mM $MgCl_2$, pH 7.4, 3 mM CaCl2). The glucose concentration was set to 11 mM for ^{13}C-labelled proteins (substituting regular glucose with D-Glucose-$^{13}C_6$). The flask was kept at 37 °C for 45 minutes and then cooled to 20 °C for 30 minutes. Protein expression was induced with 100 µM IPTG. Bacteria were pelleted 10-14 hours after induction, and frozen at –80 °C. For expression in 100% D_2O, an initial pre-culture from glycerol stock was grown in 5 ml 2xYT/Kanamycin media prepared in D_2O. The minimal medium for deuterated protein (^2H-MM) production was modified by using D_2O, anhydrous reagents and reagents dissolved in D_2O (exception: H_2O in HCl which is used for $FeCl_3$ solution preparation); glucose was not deuterated. In a typical procedure, the pre-culture was transferred to 500 ml ^2H-MM and grown to an OD of ~1.5. After centrifugation, the pellet was carefully resuspended in 2l ^2H-MM, distributed between 4 2l-flasks. At OD 0.5-0.6, the incubation temperature was lowered from 37 °C to 20 °C, followed by induction (100 µM IPTG) 30 minutes later. Bacteria were pelleted after 18-24 hours and kept frozen -80 °C.

All purification steps were performed at 4 °C. 20 ml lysis buffer (25 mM TrisCl, pH 8.0, 250 mM NaCl, 40 mM imidazole, 10 mM benzamidine, 1 mM $MgCl_2$, 3 mM $CaCl_2$, lysozyme) and one tablet cOmplete, EDTA-free protease inhibitor tablet (Roche) were added to frozen pellets in large diameter centrifugation vessels (1000 ml culture per vessel). The vessels were placed on a shaker (120 rpm) for 30 minutes. A small amount of powdered DNAse I (bovine pancrease) was then added before shaking for

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

another 30 min and centrifuging for 30 minutes (3210 g) to remove the bulk of larger cellular remnants. The supernatant was used for further purification using the AKTA Purifier system (GE Healthcare). A HisTrap crude (GE Healthcare) column and His buffers A (25 mM TrisCl, pH 8.0, 500 mM NaCl, 40 mM imidazole, 1 mM TCEP, 3 mM $CaCl_2$) and B (25 mM TrisCl, pH=8.0, 500 mM NaCl, 500 mM imidazole, 1 mM TCEP, 3 mM $CaCl_2$) were used for the first purification step. Importantly to remove impurities and reduce degradation, the column was washed with at least 300 ml His A prior to elution. The SUMO tag was cleaved by adding 100 - 500 µl of 1 mg/ml ULP1 protease for at least 20 hours, with the ULP1 concentration adjusted to achieve near complete cleavage. This was followed by at least two 3-5 hour dialyses (MW 10 kDa tubing) steps against His Chelating Buffer (25 mM TrisCl, pH=8.0, 250 mM NaCl, 20 mM imidazole, 1 mM TCEP, 3 mM $CaCl_2$) with 1 mM PMSF (stock: 0.2M in isopropanol). For the second purification step, the sample was run over a HisTrap Chelating column (GE Healthcare) to collect the tag-free flow-through. The collected sample was dialyzed (MW 10 kDa tubing) against Low Salt buffer 2x for at least 3-5 hours (25 mM TrisCl, pH 8.0, 1 mM TCEP, 3 mM $CaCl_2$). Size exclusion chromatography used a Hiload 16/60 Superdex 76 PG (Fisher Scientific) column. The sample buffer used for gel filtration is composed of a pressure resistant mixture [26] of 10 mM TRIS/tricarballylate (pH = 7, 6.905 Tris, 3.095 mM tricarballylate), 150 mM NaCl, 10% D_2O, 3 mM $CaCl_2$, followed by spin concentration using Amicon Ultra 15 Centrifugal Filters (10k kDa). Final samples contained 0.02 % (w/v) NaN_3 and a protease inhibitor mix (stock: 100x in 10% DMSO, end concentrations 0.5 mM AEBSF, 10 µM E-64, 2 mM Benzamidine; a generous gift from Eric Greene, Columbia University).

2 NMR data acquisition

2.1 general aspects

NMR data were acquired with Bruker 500-, 600-, 700-, 800- and 900-MHz spectrometers with triple-resonance z-gradient cryogenic probes (600 MHz at Columbia University, all others at New York Structural Biology Center). DSS either in the sample or separately in sample buffer was used to confirm correct temperature calibration.

2.2 2D spectra

For most TROSY-selected experiments in this work, an experimental TROSY-selected experimental framework corresponding to Igumenova et al. was used [27]. In order to enable a short recycle delay of 800 µs, the experiment was performed in a BEST-manner with an PC9,4,90 (first π/2),

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Eburp2, time-reversed Eburp2 π/2, and Rsnob π pulses [28]. 2D ^{15}N-HSQC experiments were recorded either as TROSY-selected, BEST-type TROSY, regular TROSY or SOFAST experiments. Note that the 2D spectrum in Fig. 2a is the projection sum of a pseudo-3D R_{ex} (see below) experiment, which accentuates random coil peaks due to prolonged decay of peaks belonging to slower tumbling folded regions.

2.3 3D backbone resonance assignment experiments

For backbone resonance assignment of C11-WT, TROSY versions of the HNCA, HNCOCA, HNCACB, HNCACO, HNCO and HNCACO were used as 3D experiments. For C11-W2AW4A, only HNCA, HNCO, and HNCACB experiments were recorded, for C11-W4A, only HNCA was recorded.

2.4 ^{15}N-R_2/R_1 ratios

To record ^{15}N-$R_{2,0}$/R_1 ratios, ^{15}N-R_1 and ^{15}N-$R_{2,0}$ (exchange-free) were recorded for an 80 µM sample at 285 K and 298 K. ^{15}N-R_1 was recorded as a TROSY experiment using flip-back pulses and a Watergate scheme [29]. Relaxation delays were 24 ms, 336 ms (2x for error determination), 656 ms, and 976 ms. Data were recorded in an interleaved manner, 180 points in ^{15}N dimension, total acquisition times: 1498 min (285 K), 967 min (298 K).

For exchange-free ^{15}N-$R_{2,0}$, the experimental TROSY-selected framework (section 2.2) was modified to include a relaxation period T. To suppress chemical exchange, a spin-lock was applied during T with a central S3E element dividing the two T/2 block. The experiment was set up in an interleaved manner, recording for each of the 200 ^{15}N points the following experiments: 4 x [3 x [T=15 ms with spin lock (power: 1 kHz); T=15 ms us with no spin lock]; 1 x [T=15 ms with spin lock (power: 1 kHz); T = 0 us]]. The 16 experiments with spin locks had the following offsets: 2 x 104.5 ppm; 2 x 108.5 ppm; 2 x 112.5 ppm; 2 x 116.5 ppm; 2 x 120.5 ppm; 2 x 124.5 ppm; 2 x 128.5 ppm; 2 x 132.5 ppm. To minimize off-resonance artifacts, any peaks shifted > 4 ppm (^{15}N) from the spin lock center was discarded. Therefore, per ^{15}N point, 12 spin-lock-free relaxation experiments for T=15 ms were obtained; 4 experiments with T=0 µs, and 4-6 spin-lock experiments with T=15 ms for any given residue. To obtain $R_{2,0}$, we calculate:

$$R_{2,0} = -\frac{1}{T}\log\left(\frac{I(SL,T)}{I(0)}\right)$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In this case, the exponential decay between T=0s and T=15 ms with spinlock had to be obtained; data with T=15 ms without spinlock were not needed. Note that the spinlock with B_1 = 1 kHz was not sufficient to suppress all exchange, probably accounting for the elevated $R_{2,0}$ values around residue Asp32.

2.5 R_{ex} experiments

Experiments to obtain R_{ex} were executed at 285 K using the scheme described for obtaining exchange-free ^{15}N-$R_{2,0}$ (section 2.4) with varying offsets and spin lock power levels. As a representative example, for each ^{15}N point interleaved experiments were recorded in the following order: SL, SL, noSL, SL, SL, noSL, SL, SL, noSL, SL, SL, noSL, SL, SL, T0, T0 (SL: with spin-lock; noSL: without spinlock; T0: TROSY), with SL offsets spread out over the ^{15}N spectral width. R_{ex} was obtained from

$$R_{ex} = -\frac{1}{T}\log\left(\frac{I(SL,T)}{I(noSL,T)}\right).$$

The following data sets were recorded for global fitting:

B_0 / MHz	conc / µM	T / ms	SL / kHz	ns
500	160	50	2.5	48
700	160	30	3.5	32
900	160	30	2.5	32
500	640	50	2.5	8
900	640	30	2.5	8
900	640	30	2.5	8

with SL: spin long power; ns: number of scans

In the experiment in the last line, interleaved off-resonance low-power T1 and off-resonance R1rho experiments were included but not used in the analysis.

Error estimation: The average of the square root of the squared standard error of the absolute intensity for each residue was obtained. This averaged error was then propagated to R_{ex} individually for each residue using Monte Carlo simulations. We found that residue-specific differences in error mostly stem from absolute intensity differences, and those are preserved in our approach.

2.6 ^{15}N-CPMG experiments

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The TROSY-selected framework used for the ^{15}N-$R_{2,0}$ and Rex experiments was modified by replacing the spin-lock block with constant-time CPMG blocks. Each half of the CPMG block, divided by an S3E-element, contains N repetitions of 4 π pulses (phases: x, x, y, -y), separated by variable τ_{cp}. The various of N and τ_{cp}, constant-time (total time T) CPMG blocks were recorded in an interleaved manner, with the objective to average heating by not sequentially recording two experiments with similary high (short τ_{cp}) or low (long τ_{cp}) heating .

The following data sets were recorded for global fitting:

B_0 / MHz	conc / µM	T / ms	ns
500	2.475	60	40
900	2.475	45	32
500	9.9	30	8
500	9.9	30	8
900	9.9	45	8

In analogy to our error estimation for R_{ex} method but having the benefit of multiple data points per curve, we obtained average absolute intensity errors for each curve collected from all τ_{cp} values for which more than one τ_{cp} point had been recorded. The error for each intensity ratio (I_T/I_0) and subsequently R_{cpmg} was then calculated using standard error propagation. The positive and negative errors were calculated separately unless the negative error calculation led to negative intensity ratios.

2.7 ^{15}N-CEST experiments

Some ^{15}N-CEST experiments were based on the Bruker library pulse sequence hsqc_cest_etf3gpsitc3d (15/09/22), which is a non-TROSY sequence with decoupling during acquisition. Alternatively, a TROSY-selected version of the experiment was run, based on the framework for ^{15}N-$R_{2,0}$ and R_{ex}. During the CEST period T, CPD decoupling was employed in the proton channel. All experiments, including various CEST offsets and reference experiments were run in an interleaved manner at 285K.

For the 160 µM sample, three non-TROSY experiments at 700 MHz with T = 400 ms for three different B1 fields, 15 Hz, 25 Hz, and 50 Hz, were recorded. For the 640 µM sample, the TROSY-experiment was used with T = 400 ms and B_1 = 35 Hz.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For further processing, ^{15}N-CEST data were trimmed to reduce the number of information-free data points. Separately, sections with frequency distant from the CEST dip were used to define reference CEST intensities and to obtain the intensity errors. CEST data points < 0.5 ppm away from the CEST dip minimum were discarded for the CEST analysis.

2.8 ^{1}H-CPMG experiments

^{1}H-CPMG experiments were performed described following the pulse sequence provided by Li et al. [5]. Due to a relatively fast degradation an of the deuterated sample which was used for the experiments, the total acquisition times per sample were of limited duration, giving large error bars. Uncertainties in the concentrations an imperfect comparability of temperature between experiment limit the recorded data to be used for qualitative purposes only. For aspects of initial error calculation, see section 2.6.

2.9 High pressure experiments

The Daedalus Xtreme-60 apparatus was used for static high pressure NMR experiments. Mineral oil was used as the buffer fluid for the pressure tubes. The standard air removal procedure was repeated three times prior to connecting the sample. For titration experiments, the experiments were started at a high pressure and then reduced stepwise. To rule out problems with equilibration, control experiments, either by repeating by changing the direction of pressure change, were performed. We found that in our NMR buffer system, pressure hardly affects the ^{1}H reference shift.

3 NMR data processing

3.1 General approach

Raw NMR data were pre-processed using customized bash scripts and NMRPipe [30]. CCPN analysis 2.5.0 [31] was used for peak assignment and extraction of peak intensities and chemical shifts. For peaks with particularly small intensities, e.g. at noise level, the peak fitting algorithm tends to overestimate peak intensities. For that reasons we decided to obtain peak intensities by adding the sum of all spectra from any pseudo-3D dataset to each individual spectrum for fitting at the preprocessing stage with NMR Pipe. The peak intensities obtained from the sum of all spectra are then subtracted from the sum of all spectra + spectrum of interest after peak fitting with CCPN analysis. Depending on the experiment type, specific data formats were established, also to include all relevant experimental conditions, to ensure smooth automated postprocessing; conversion bash scripts were used to avoid any errors and reveal/fix any inconsistencies in the peak lists obtained from CCPN Analyis. Python was used for any subsequent

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

data processing. Errors for data points were calculated as described in the NMR data acquisition section 2.

3.2 spin system objects, data sets

In our python code framework, each spin systems is associated with a protein sequence position, protein name, state (A: folded; B: unfolded/RC; C: dimer; D,E…: other). Each of the spin system objects contains a list of data set objects which behave like dictionaries. These dictionaries can contain a variety of entries, typically several experimental conditions, intensity data and/or chemical shifts. During subsequent data processing, the number of entries can grow, for example due to a temperature correction of the chemical shifts. There can also be created non-experimental datasets generated by fitting procedures. In this way, the results of a variety of fitting procedures are stored in any spin system object.

3.3 multi-site fitting process

We used Python scripts for data processing. To enable flexibility in data and parameter choice we defined dedicated property dictionary objects, in which any property relevant in the fitting process can be stored. Specifically, fitted kinetic parameters can depend on the B_0 static magnetic field, experimental temperature, site (in a kinetic scheme), protein concentration, B_1 irradiating magnetic field, experimental type, type of TROSY experiment, residue/spin system in question. Each of the property categories constitutes a dimension for the parameter object, which contains all parameters for all properties, but of which there can exist multiple copies when the fitting process is being repeated (for example with different fitting conditions).

We selected the following approximations for fitting. For the global fit, ^{15}N-CEST data were converted to $R_{1\rho}$-type data as described elsewhere [23], using $R_{1\rho}$ multi-site approximations for initial calculations [32]. CPMG data were fitted with the $Exp_0Log_2\lambda_2$ approximation for the triangular state [33]. R_{ex} data were fit using exact calculations.

Multiple initial tests with wide boundary conditions and inspection of the relaxation dispersion datasets corroborate to use at least a three-site model, including a fast and a slow exchange process, to be able to fit the data. We find that the appropriate scheme fast exchange between sites A and B and slow exchange between sites A and C, with A being the major site. The initial fit gave broad result ranges for parameters, specifically for all Δω values and k_{ex} which are known to be highly correlated. We tested a variety of Δω restraints. We also found that the Δω of different residues were highly correlated for Δω >>

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

0 (e.g. $\Delta\omega_{AC}$ of residues 38 and 14). Therefore, fixing any of the $\Delta\omega >> 0$, of which there are a total of 2*(number of residues), greatly shrinks the parameter space of fitting results. Initially, we attempted to set $\Delta\omega_{AC}$ to known chemical shift perturbations: The chemical shift position for the main state A was set to be the position of the recorded peak which is justified by the fact that the corresponding exchange process is generally in the slow exchange regime. Chemical shift for position C was set in the following ways: (a) chemical shift of other states or mutants, specifically the unfolded state, the folded state of the W2A mutant, the folded state of the W4A mutant, the folded state of the W2AW4A mutant, and the dimer state; (b) The sign of the chemical shift was guessed from CEST curve if a clear asymmetry was visible. For W2A, there were no conflicts between (a) and (b). The following residues were restrained based on W2A shifts: 30, 32, 38, 43, 45, 53, 54, 55, 73, 77, 78, 86. For residues 38,73,77 and 86, the sign of the chemical shift was inferred directly from the shape of the CEST curve. $\Delta\omega_{AC}$ for residues 37 and 50 was set freely.

However, for residues Gly43 and Gly55, very large, unrealistic ω_B emerge under any scenario leading to reasonable χ^2 for these residues. Close inspection of the relaxation dispersion curves suggests partial quenching of a separate very fast exchange process which is either unique to these residues or equivalent to the otherwise very fast folded-unfolded process, which might become detectable due to the unique contribution of exchange processes between the main site A and the minor sites B and C.

The main result uses the following chemical shifts as restraints: site A: WT-Cad11; site B: folded W2AW4A-Cad11 and unfolded W2AW4A-Cad11 shifts as boundaries, but only for residues which display "clean" fast exchange behavior (BC and FG loops); site C: W2A-Cad11. With these restraints, χ^2 (1.214-1.247) is not significantly increased with respect to no restraints. Exchange between minor sites B and C is possible but does not change χ^2. Examples for other non-fitting restraints which do give higher χ^2 include setting the following chemical shifts: site C to W4A-Cad11, site C to WT-dimer, site B to W2A-Cad11 or site B (generally or for the specified regions) to unf-W2AW4A-Cad11. Setting site B to fold-W2AW4A-Cad11 for loop1 and 2 increases χ^2 moderately. We tested the outlined $\Delta\omega$ restraints with all BC, DE and FG residues, including data with almost flat relaxation profiles or with large uncertainties (***Suppl. Fig. 12***): We used all known W2A and W2AW4A chemical shifts for setting $\Delta\omega$ restraints and compared our obtained model (with fixed $k_{12+}k_{21}$, $k_{13+}k_{31}$, p_b, p_c and $\Delta\omega$ from known W2A and W2AW4A chemical shifts) to unrestrained runs for each residue. Leaving only $\Delta\omega$ unrestrained did not improve χ^2. Leaving $k_{12}+k_{21}$, $k_{13}+k_{31}$, p_b and p_c free for each residue gives only very slightly lower χ^2 than restraining it globally. Allowing chemical exchange between the minor sites B and C does not change χ^2.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

After obtaining a global fit solution, the error was estimated using permutation tests. For that purpose, for each of the data sets (individual CEST, CPMG curves, all R_{ex} per residue), residuals $\delta_i = y_{i,fit} - y_{i,exp}$ were calculated. The lists of residuals were randomized to create new datasets with $\delta_{j,new} = y_{j,fit} + \delta_i$. The new datasets were than again globally fit to obtain an alternative set of parameters. Multiple repetitions of this process were used to produce parameter ranges of 1σ uncertainty.

3.4 mapping chemical exchange to the structure by evaluating spectra recorded at different B_0 and temperatures

The methods shown here aims to include relaxation processes encoded in the 15N and 1H dimension and peaks with intensities too low for RD analysis. We estimated the degree of slow-intermediate exchange between the main state and the W2A-like (partially strand exposed) state directly from combining knowledge from the RD analysis with peak intensities recorded at different conditions. Fast exchange displays a pronounced temperature and field dependency of peak intensities. Fast exchange at 298K is essentially invisible in RD analysis. For peaks with slow-intermediate exchange, we also find a different field dependency of peak intensities. We therefore calculated the intensity ratio

$$\frac{I_{900MHz}^{285K}}{I_{500MHz}^{285K}} \Bigg/ \frac{I_{900MHz}^{298K}}{I_{500MHz}^{298K}}$$

from $^{1}H,^{15}N$-TROSY-HSQC spectra and plotted it on the structure. A large ratio maps to a large $\Delta\omega_{AC}$, which corresponds to a site with major exchange between a major strand-bound and a sparsely populated partially strand-exposed (W2A-like) state.

3.5 obtaining random coil populations of very fast exchanging residues

The relative population of random coil is, by definition, 100% in the RC state. While it cannot be ruled out that the fraction of remnant random-coil state in the WT-dimer state is >> 0%, the fact that its main peak position is well in the folded region of the spectrum suggests that this fraction will not be large. For purposes of comparing relative folded/RC populations for WT-monomer and mutant states, we assume WT-dimer to be the reference state. We observe that, while the various mutants generally follow the order mentioned above, the degree of disorder varies from residue to residue.

3.6 dependency of chemical shifts and intensities on high pressure

The processed pressure dependency data for intensities and chemical shifts (^{1}H, ^{15}N) included data for multiple residues, two states (folded/unfolded) and two temperatures (285K, 298K). The pressure

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

profiles were not reach and the plateau values for the most unfolded state was not reached. For that reason, we attempted to extract the characteristics of the curves using spline functions, fitting intensities and chemical shifts at once, but individually for each residue. The first-order two-point spline function did not give satisfactory fits in many cases. We found that either second-order two-point or first-order three-point spline functions generally resulted in good fits, except for the folded peaks for some residues in the FG loop and for residue Ser54 in the DE loop. For the main analysis we refer to the first order spline fit with three points (two straight lines) which returns a critical pressure for each fit. Intensities were scaled to 1 bar (folded peaks) or 1.72 kbar (unfolded peaks).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Brasch, J., et al., *Thinking outside the cell: how cadherins drive adhesion.* Trends Cell Biol, 2012. **22**(6): p. 299-310.
2. Brasch, J., et al., *Homophilic and Heterophilic Interactions of Type II Cadherins Identify Specificity Groups Underlying Cell-Adhesive Behavior.* Cell Rep, 2018. **23**(6): p. 1840-1852.
3. Patel, S.D., et al., *Type II cadherin ectodomain structures: implications for classical cadherin specificity.* Cell, 2006. **124**(6): p. 1255-68.
4. Harrison, O.J., et al., *Two-step adhesive binding by classical cadherins.* Nat Struct Mol Biol, 2010. **17**(3): p. 348-57.
5. Li, Y., et al., *Mechanism of E-cadherin dimerization probed by NMR relaxation dispersion.* Proc Natl Acad Sci U S A, 2013. **110**(41): p. 16462-7.
6. Shapiro, L., et al., *Structural basis of cell-cell adhesion by cadherins.* Nature, 1995. **374**(6520): p. 327-337.
7. Miloushev, V.Z., et al., *Dynamic properties of a type II cadherin adhesive domain: implications for the mechanism of strand-swapping of classical cadherins.* Structure, 2008. **16**(8): p. 1195-205.
8. Sivasankar, S., et al., *Characterizing the initial encounter complex in cadherin adhesion.* Structure, 2009. **17**(8): p. 1075-81.
9. Moschen, T. and M. Tollinger, *A kinetic study of domain swapping of Protein L.* Phys Chem Chem Phys, 2014. **16**(14): p. 6383-90.
10. Rousseau, F., J.W. Schymkowitz, and L.S. Itzhaki, *The unfolding story of three-dimensional domain swapping.* Structure, 2003. **11**(3): p. 243-51.
11. Byeon, I.J., J.M. Louis, and A.M. Gronenborn, *A captured folding intermediate involved in dimerization and domain-swapping of GB1.* J Mol Biol, 2004. **340**(3): p. 615-25.
12. Esposito, L. and V. Daggett, *Insight into ribonuclease A domain swapping by molecular dynamics unfolding simulations.* Biochemistry, 2005. **44**(9): p. 3358-68.
13. Liu, Y., et al., *A domain-swapped RNase A dimer with implications for amyloid formation.* Nat Struct Biol, 2001. **8**(3): p. 211-4.
14. Liu, L., et al., *Domain swapping proceeds via complete unfolding: a 19F- and 1H-NMR study of the Cyanovirin-N protein.* J Am Chem Soc, 2012. **134**(9): p. 4229-35.
15. Daoud, M., et al., *Solutions of Flexible Polymers. Neutron Experiments and Interpretation.* Macromolecules, 1975. **8**(6): p. 804-818.
16. Gronenborn, A.M., *Protein acrobatics in pairs--dimerization via domain swapping.* Curr Opin Struct Biol, 2009. **19**(1): p. 39-49.
17. Spadaccini, R., et al., *Mechanism of 3D domain swapping in bovine seminal ribonuclease.* FEBS J, 2014. **281**(3): p. 842-50.
18. Kang, X., et al., *Foldon unfolding mediates the interconversion between M(pro)-C monomer and 3D domain-swapped dimer.* Proc Natl Acad Sci U S A, 2012. **109**(37): p. 14900-5.
19. Liu, Z. and Y. Huang, *Evidences for the unfolding mechanism of three-dimensional domain swapping.* Protein Sci, 2013. **22**(3): p. 280-6.
20. Yang, S., et al., *Domain swapping is a consequence of minimal frustration.* Proc Natl Acad Sci U S A, 2004. **101**(38): p. 13786-91.
21. Chang, S.K., Z. Gu, and M.B. Brenner, *Fibroblast-like synoviocytes in inflammatory arthritis pathology: the emerging role of cadherin-11.* Immunol Rev, 2010. **233**(1): p. 256-66.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

22. Sfikakis, P.P., N.I. Vlachogiannis, and P.F. Christopoulos, *Cadherin-11 as a therapeutic target in chronic, inflammatory rheumatic diseases.* Clin Immunol, 2017. **176**: p. 107-113.
23. Palmer, A.G., 3rd and H. Koss, *Chemical Exchange.* Methods Enzymol, 2019. **615**: p. 177-236.
24. Vendome, J., et al., *Molecular design principles underlying beta-strand swapping in the adhesive dimerization of cadherins.* Nat Struct Mol Biol, 2011. **18**(6): p. 693-700.
25. Murray, V., et al., *A novel bacterial expression method with optimized parameters for very high yield production of triple-labeled proteins.* Methods Mol Biol, 2012. **831**: p. 1-18.
26. Quinlan, R.J. and G.D. Reinhart, *Baroresistant buffer mixtures for biochemical analyses.* Anal Biochem, 2005. **341**(1): p. 69-76.
27. Igumenova, T.I. and A.G. Palmer, *Off-resonance TROSY-selected R 1rho experiment with improved sensitivity for medium- and high-molecular-weight proteins.* J Am Chem Soc, 2006. **128**(25): p. 8110-1.
28. Lescop, E., T. Kern, and B. Brutscher, *Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: Example of longitudinal-relaxation-enhanced BEST-type 1H–15N correlation experiments.* Journal of Magnetic Resonance, 2010. **203**(1): p. 190-198.
29. Zhu, G., et al., *Protein dynamics measurements by TROSY-based NMR experiments.* J Magn Reson, 2000. **143**(2): p. 423-6.
30. Delaglio, F., et al., *NMRPipe: A multidimensional spectral processing system based on UNIX pipes.* Journal of Biomolecular NMR, 1995. **6**(3): p. 277-293.
31. Vranken, W.F., et al., *The CCPN data model for NMR spectroscopy: development of a software pipeline.* Proteins, 2005. **59**(4): p. 687-96.
32. Koss, H., M. Rance, and A.G. Palmer, 3rd, *General expressions for R1rho relaxation for N-site chemical exchange and the special case of linear chains.* J Magn Reson, 2017. **274**: p. 36-45.
33. Koss, H., M. Rance, and A.G. Palmer, 3rd, *General Expressions for Carr-Purcell-Meiboom-Gill Relaxation Dispersion for N-Site Chemical Exchange.* Biochemistry, 2018. **57**(31): p. 4753-4763.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

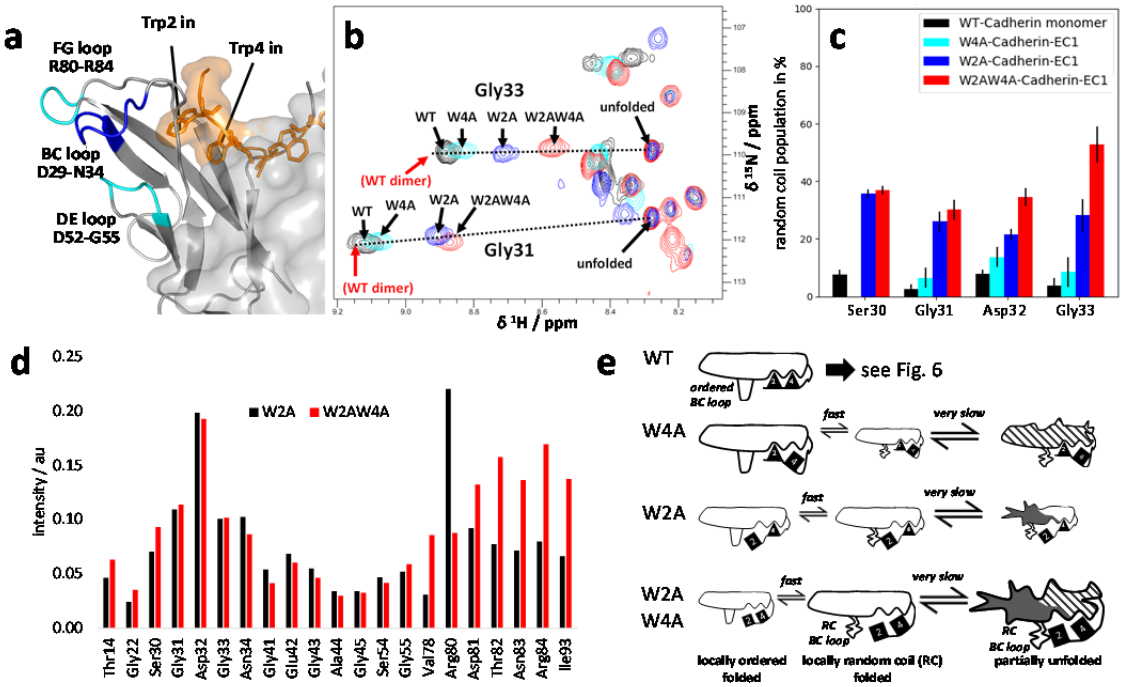

Fig. 1: Chemical exchange in Trp2 and Trp4 mutants. a: Structural features of Cad-11 EC1. Orange: N-terminal strand with Trp2,Trp4 of homodimerized Cad-11 EC1, inserted in the intermolecular binding pocket. The BC, DE, and FG loops, which display unique dynamic features without confounding secondary chemical shift perturbations, are indicated. **b: Superposed ^{1}H,^{15}N-HSQC spectra for WT, W2A, W4A, and W2AW4A mutants of Cadherin-11-EC1.** The position of the WT dimer peaks (not visible in this dilute sample) is indicated. Dotted lines are drawn from the dimer resonance to the RC resonance for each amino acid residue. The random coil (RC) peak position is common to all constructs, but intensities vary greatly between mutants. The RC population for different constructs varies between residues (see also Suppl. Fig. 4). **c: Semi-quantitative estimation of RC populations from projections on the ^{1}H,^{15}N-vector.** The end points for the RC population destination are ordered state and RC state. The dimer peak was used as a reference for the ordered state. The true RC populations various states might be slightly higher than indicated, with preserved relative scaling for all mutants belong to one residue. **d: Normalized unfolded peak intensities of W2A and W2AW4A mutants.** The data suggest similar partially unfolded structures except for the region encompassing the FG loop. Intensities were scaled to the average intensity of unfolded peaks for all residues in sequence up until residue 55. **e: Fast and very slow exchange based on the mutant spectra.** Sizes of cartoons represent populations (not to scale). Positions and status of Trp2, Trp4, and BC loop are indicated. The various patterns and shading in the partially unfolded structure represents common substructures among partially unfolded mutant domains, also used in Fig. 3f and Fig. 6.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

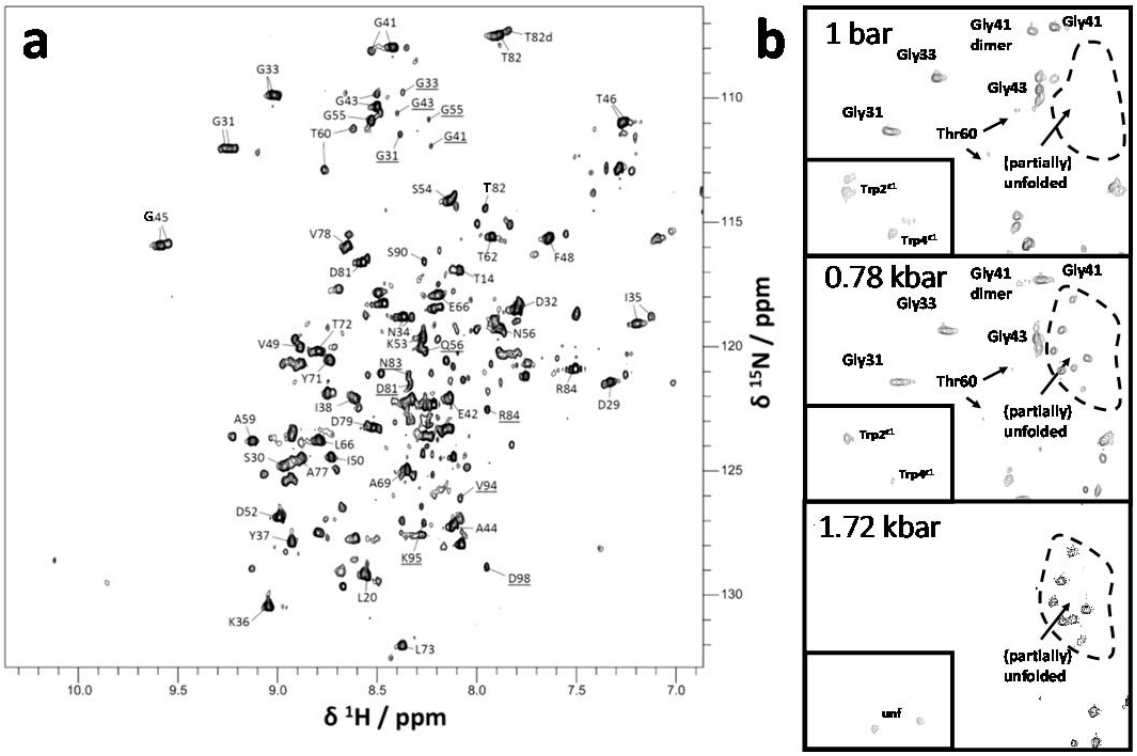

Fig. 2: ^{1}H,^{15}N- TROSY-HSQC spectra of Cad11-EC1-WT at different pressures. a: ^{1}H-^{15}N correlation spectra recorded at 900 MHz, 160 mM, and 298 K, and 1 bar pressure. Duplicate assignments mostly refer to monomer/dimer pairs. Underlined peaks refer to random coil peaks, which are hardly detectable in WT samples. The two marked regions refer to the spectral regions in the right panel. **b: Pressure series recorded at 800 MHz, 160 µM and 285 K.** The pressures are shown in the figure. The corresponding areas in the full ^{1}H,^{15}N spectrum are marked in the left panel; see also Suppl. Fig. 6.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

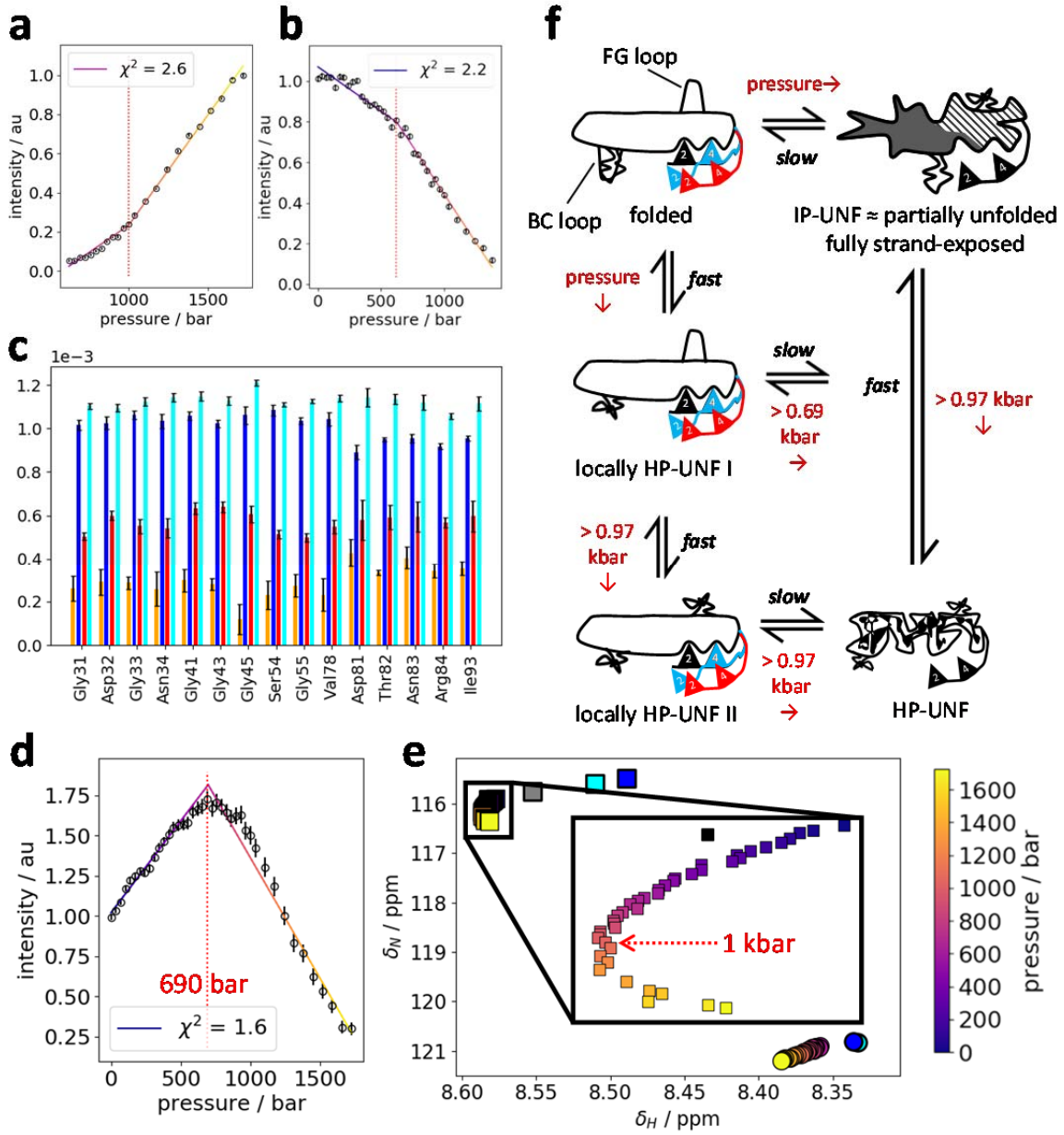

Fig. 3: Analysis of pressure dependence of peak intensities. a,b: Pressure dependency of Gly33 peak intensities at 285K. A first order spline with three points (two straight lines) was used to characterize each of the three intensity profiles; panel a shows the intensity of the RC (unfolded) peak, panel b the intensity of the folded peak. The dotted red line indicates the pressure of the central critical pressure for the three-point first order spline fit. For more examples see Suppl. Fig. 8. χ^2 is shown in the figure legend. **c: Fitted slopes for intensity dependency on pressure of unfolded peaks at 283 K and 298 K.** Three-point first order spline (right panel) fits give two separate slopes: The low-pressure slope for 283K (red)

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and 298K (orange), and the high-pressure slopes for 283K (cyan) and 298K (blue). **d,e: Pressure dependency of folded Asp81 peak intensities (panel d) and chemical shifts (panel e) at 285K.** The intensities in panel d were fitted in the same manner as described in panel a. The color of the calculated curve indicates pressure. The same pressure scale is used for intensities in panel e, alongside a color bar (right) translating colors into pressures. Other symbols in panel e: squares – folded peaks; circles – unfolded (RC) peaks; blue: W2AW4A mutant peak; cyan: W2A mutant peak; grey: W4A mutant peak: black: WT peak for comparison with mutants. The critical pressures for intensities (panel d) and chemical shifts (panel e) differ, indicating a three-step unfolding process: first, general decline of the population (slow exchange type), then local unfolding. Resonance frequency shifts in ^{1}H,^{15}N-TROSY spectra recorded at pressures from 1 bar to 1.72 kbar result from population changes between conformational states in fast exchange. The folded peak for Asp81 shifts away from the unfolded peak at low pressures and "changes direction" and moves towards the unfolded state only above ≈ 1 kbar (inset). **f: Scheme summarizing the effects of pressure on Cad11-EC1 conformations.** The depictions of the various states for Trp2, Trp4 and folded FG loop are schematic (black – strand-bound; cyan – partially strand exposed; red – fully strand-exposed; loops depicted as folded, ambient-pressure unfolded and high-pressure unfolded). The shade for the partially unfolded states on the right is corresponding to the shade in Fig. 1e and Fig. 6. For more pressure-related data, see Suppl. Figs. 7-9.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

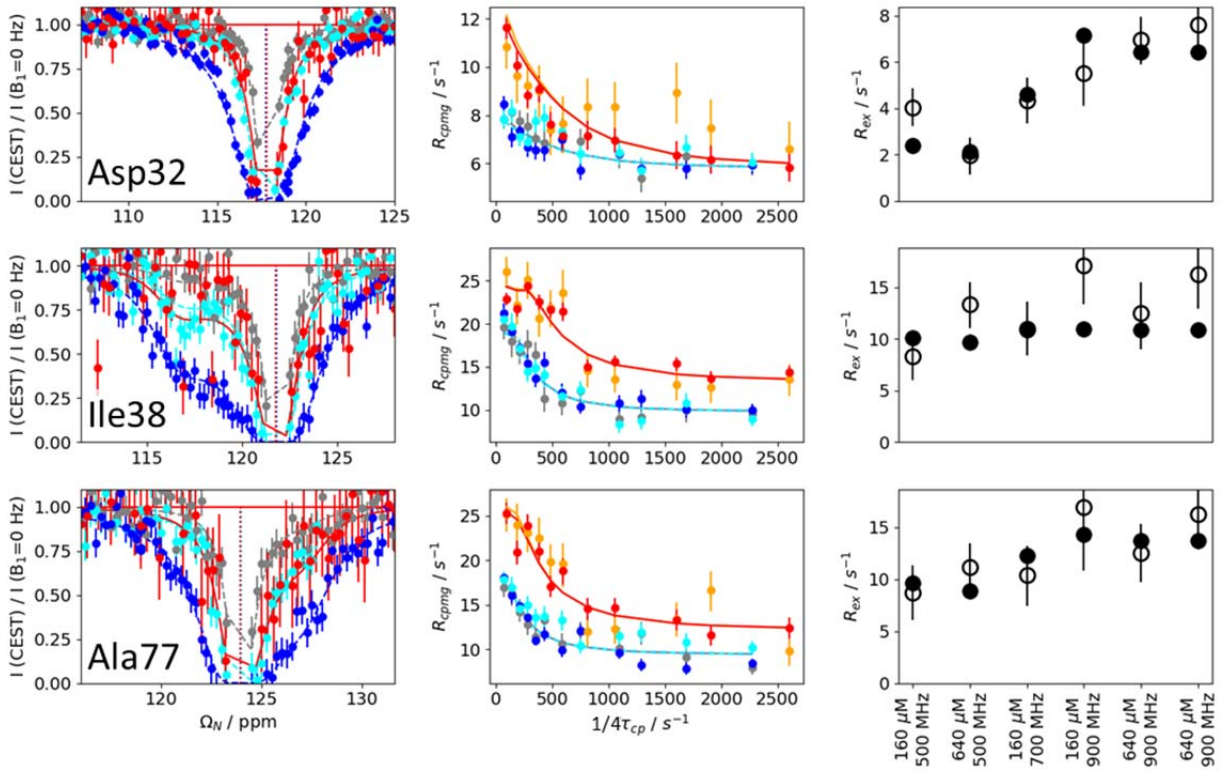

Fig. 4: ^{15}N-relaxation dispersion reveals three-site exchange. Three of the 12 globally fitted residues have been selected to exemplify various type of chemical exchange observed within Cadherin-11-EC1 domain: Val38: mostly slow-intermediate exchange; Ser32: mostly fast-intermediate exchange; Ala77: combination of slow-intermediate and fast-intermediate exchange. Left panels: ^{15}N-CEST at at c = 640 µM, B_0 = 800 MHz, B_1 = 35 Hz (red) or at c = 160 µM and B_0=700 MHz (grey – B_1 = 15 Hz, cyan – B_1 = 15 Hz, blue – B_1 = 50 Hz); center panels: ^{15}N-CPMG at c = 160 µM (grey, orange), c = 640 µM (blue, cyan, red), B_0 = 500 MHz (grey, cyan, blue), B_0 = 900 MHz (orange, red); right panels: ^{15}N-high power spin-lock experiments for R_{ex} determination, with conditions listed along the x axis with experimental (open circles) and calculated (filled circles) values. For the global fit, ^{15}N-CEST data were converted to $R_{1\rho}$-type data as described elsewhere [23], using $R_{1\rho}$ multi-site approximations for initial calculations [32]. CPMG data were fitted with the $Exp_0Log_2\lambda_2$ approximation for the triangular state [33]. For all data, see Suppl. Fig. 9.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

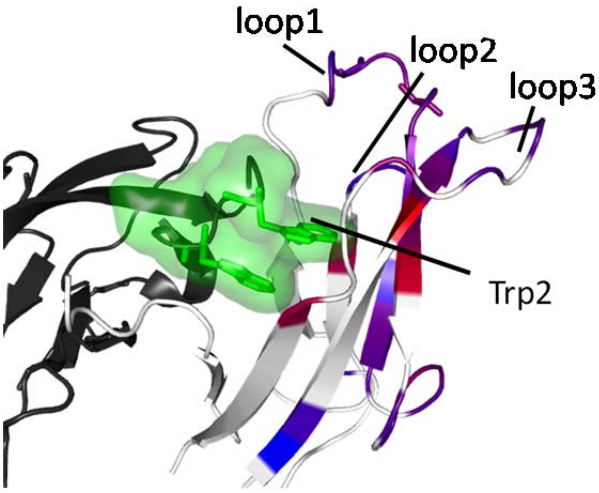

Fig 5: The Trp2-like minor state. The peak intensity ratio $I^{298K,500MHz} * I^{285K,900MHz} / I^{298K,900MHz} * I^{285K,500MHz}$ is projected on the structure (blue – violet – red). The result confirms the expectation from relaxation dispersion analysis that the ratio encodes the chemical shift perturbation between WT and the minor W2A-like state.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.21.423864; this version posted December 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

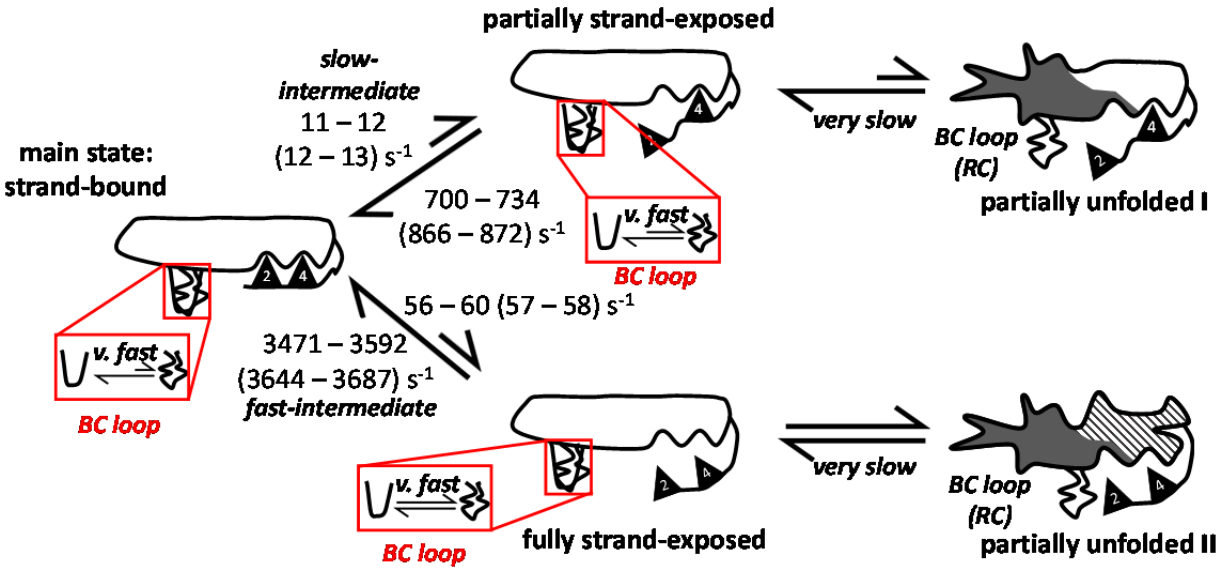

Fig. 6: Chemical exchange in Cad11-EC1 WT. Scheme illustrating chemical exchange at various dynamic levels. The local very fast equilibria for BC loop are shown in red insets, with the equilibrium arrow indicating general propensity towards the ordered or the RC state. Possible very fast exchange for FG loop is indicated but not specifically shown by a blue box. Slow-intermediate and fast-intermediate exchange have been established by relaxation dispersion. Resulting rate constant ranges for low and high concentrations are shown outside or inside brackets, respectively.