

# Disordered Regions Flanking the Binding Interface Modulate Affinity between CBP and NCOA

Elin Karlsson<sup>1†</sup>, Jan Schnatwinkel<sup>1,2†</sup>, Cristina Paissoni<sup>3</sup>, Eva Andersson<sup>1</sup>, Christian Herrmann<sup>2</sup>, Carlo Camilloni<sup>3\*</sup> and Per Jemth<sup>1\*</sup>

**1 - Department of Medical Biochemistry and Microbiology, Uppsala University, BMC Box 582, SE-75123 Uppsala, Sweden**

**2 - Department of Physical Chemistry I, Faculty of Chemistry and Biochemistry, Ruhr University Bochum, D-44780 Bochum, Germany**

**3 - Dipartimento di Bioscienze, Università degli Studi di Milano, 20133 Milano, Italy**

**Correspondence to Carlo Camilloni and Per Jemth:** [carlo.camilloni@unimi.it](mailto:carlo.camilloni@unimi.it) (C. Camilloni), [per.jemth@imbim.uu.se](mailto:per.jemth@imbim.uu.se) (P. Jemth), [@CarloCamilloni](https://twitter.com/CarloCamilloni) (C. Camilloni), [@JemthLab](https://twitter.com/JemthLab) (P. Jemth)  
<https://doi.org/10.1016/j.jmb.2022.167643>

**Edited by Richard W. Kriwacki**

## Abstract

Recognition motifs that mediate protein–protein interactions are usually embedded within longer intrinsically disordered regions. While binding interfaces involving the recognition motif in such interactions are well studied, less is known about the role of disordered regions flanking the motifs. The interaction between the transcriptional co-activators NCOA3 (ACTR) and CBP is mediated by coupled binding and folding of the two domains CID and NCBD. Here, we used circular dichroism and kinetics to directly quantify the contribution of the adjacent flanking regions of CID to its interaction with NCBD. Using N- and C-terminal combinatorial variants we found that the flanking regions promote binding in an additive fashion while retaining a large degree of disorder in the complex. Experiments at different ionic strengths demonstrated that the increase in affinity is not mediated by electrostatic interactions from the flanking regions. Instead, site-directed mutagenesis and molecular dynamics simulations suggest that binding is promoted by short-lived non-specific hydrophobic contacts between the flanking regions and NCBD. Our findings are consistent with highly frustrated interactions outside of the canonical binding interface resulting in a slightly energetically favorable fuzzy binding. Modulation of affinity via flanking regions could represent a general mechanism for functional regulation by intrinsically disordered protein regions.

© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

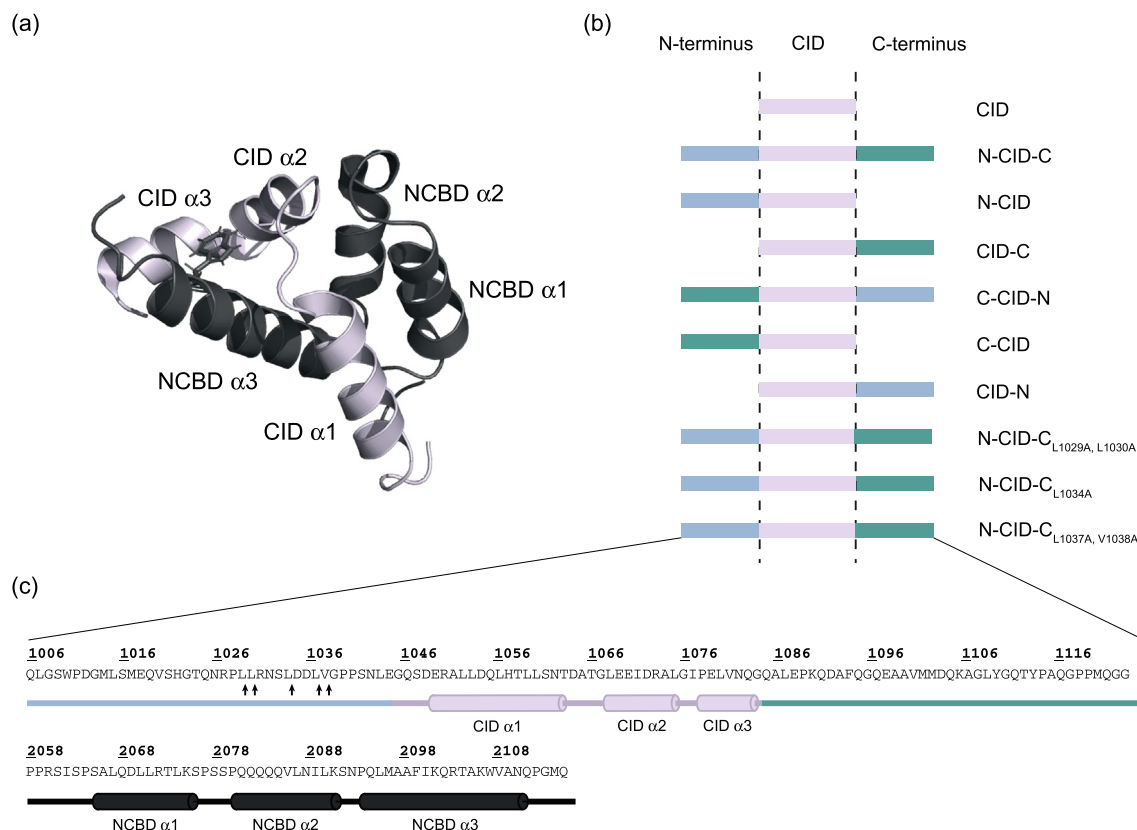
A common feature of proteins involved in cellular signaling is that they have regions lacking a well defined tertiary structure.<sup>1</sup> Such intrinsically disordered proteins (IDPs) are abundant in eukaryotic cells and especially in protein–protein interaction networks, where they may act as hubs with the ability to interact with a large number of different binding partners.<sup>2</sup> IDPs interact with their partners via binding motifs of variable size, ranging from a few to

over 20 amino acid residues.<sup>3–4</sup> Upon binding, these motifs often adopt a well folded structure, stabilized by intermolecular interactions with their binding partners, but motifs can also display structural heterogeneity in the bound state.<sup>5–6</sup> The structural plasticity and overall heterogeneity of IDPs are believed to facilitate the evolutionary optimization of binding interfaces such that the affinity is fit for the biological function.<sup>7–8</sup> However, the interaction motifs are commonly found in longer disordered regions in the protein, and little is known about the

role of the regions flanking the binding motif, and how they contribute to modulating affinity.<sup>9–11</sup> As the folding and binding reactions of IDPs are commonly characterized by shallow energy landscapes<sup>12</sup> they are particularly susceptible to local changes in chemical environment and interactions made by flanking regions therefore emerge as a possible mechanism for adaptation of functional affinity.

The nuclear coactivator binding domain (NCBD) of the CREB binding protein (CBP) interacts with the CBP interaction domain (CID) of the p160 transcriptional co-activator NCOA3 (also called ACTR).<sup>13</sup> In the free state NCBD is very dynamic and displays molten globule like properties.<sup>14–17</sup> CID on the other hand is a 42 amino acid long highly disordered sequence with rather low helical propensity.<sup>18</sup> It contains a nearly 40 residues long interaction motif composed of shorter hydrophobic leucine rich motifs of the type  $\Phi\Phi\text{XX}\Phi$  and  $\Phi\text{XX}\Phi\Phi$ , where  $\Phi$  represents a bulky hydrophobic residue and X any residue. Upon binding to NCBD, CID adopts an extended  $\alpha$ -helical conformation ( $K_D \sim 0.1\text{--}0.3 \mu\text{M}$ ). The structure of the complex

between NCBD and CID (Figure 1) was the first reported example of a mutual synergistic folding and binding reaction of two IDPs<sup>19–20</sup> and CID/NCBD has since served as a model system for coupled binding and folding. Thus, the binding mechanism of NCBD and CID has been extensively studied with experimental<sup>21–25</sup> and computational<sup>26–28</sup> methods and involves a malleable pathway that can be understood in terms of templated folding.<sup>29</sup> Despite the complexity of the binding reaction, the affinity of the interaction can be studied using kinetics under apparent two-state conditions with great accuracy and precision in the parameters. Here we use the CID/NCBD system to address the fundamental question of the role played by flanking regions in protein interactions. We designed combinatorial variants of CID with the flanking sequences either in their native locations or in swapped positions (Figure 1) and analyzed their effect on binding. Our results demonstrate how binding plasticity involving disordered flanking regions of IDPs may underlie tuning of their functional affinity in interaction networks.



**Figure 1. Sequences and structures of CID and NCBD.** (a) Structure of the CID/NCBD complex. NCBD is shown in dark grey and CID in light purple/grey. Trp2108 is the fluorescent probe used in stopped flow experiments, and is highlighted as a stick model. The graphic was generated with PyMol from PDB entry 1kbh. (b) The CID constructs used in the experiments. (c) Sequence of NCBD<sub>Y2108W</sub> and N-CID-C with the helical regions marked. The arrows indicate where the point mutations are located in the N-terminal region of N-CID-C.

## Results

### Variants of CID with different lengths were used to test the role of flanking regions

In the initial work on the CID/NCBD interaction a 71 amino acid long fragment from NCOA3 corresponding to the CID domain (residues 1,018 to 1,088 from NCOA3) was shown to mediate the interaction with NCBD.<sup>13</sup> Using this fragment, Wright and coworkers were able to solve the NMR structure of the CID/NCBD complex and show that residues 1,018 to 1,045 and 1,084 to 1,088 were completely unstructured, while 1,046–1,083 formed the binding interface with NCBD.<sup>19</sup> The same CID construct (residues 1,018 to 1,088) has been used in kinetic studies,<sup>22,24–25</sup> whereas recent single molecule studies conducted by Schuler and coworkers employed a CID variant comprising residues 1,023 to 1,093.<sup>17,21</sup> Kinetics and affinity data from these studies cannot be directly compared, as they were conducted not only with different constructs, but also under different buffer conditions and temperatures. However, we found during evolutionary studies on CID/NCBD<sup>30</sup> that a shorter CID construct (residues 1045 to 1086), corresponding to the evolutionarily well-conserved NCBD-binding interface, displayed a slightly lower affinity for NCBD as compared to CID<sub>1,018-1,088</sub> when experiments were performed under identical conditions. Inspired by this observation, we set out to systematically map the role of flanking regions in the interaction between NCBD and CID from NCOA3. In the present study we chose residues 1045–1086 as the minimal interaction domain, henceforth referred to as CID. The sequence of CID and both of its adjacent amino-terminal and carboxy-terminal flanking regions (each 39 amino acids in length) is called N-CID-C. To systematically test the effect of these regions on binding we designed variants of CID including combinations of N-terminal and C-terminal flanking regions, in native as well as in swapped positions: N-CID, CID-C, C-CID-N, C-CID and CID-N (Figure 1). Of note, the N-terminal flanking region of CID contains a putative leucine-rich motif (residues 1,029–1,038) that was included in the original construct, but was not resolved in the NMR structure.<sup>12</sup> This leucine-rich motif has a high similarity to the  $\Phi$ XX $\Phi$ XX $\Phi$  motif in the  $\alpha_1$  helix of CID in the complex (Figure S1). To investigate the role of this motif in the N-terminus we generated three mutants of N-CID-C: N-CID-C<sub>L1029A L1030A</sub>, N-CID-C<sub>L1034A</sub> and N-CID-C<sub>L1037A V1038A</sub>.

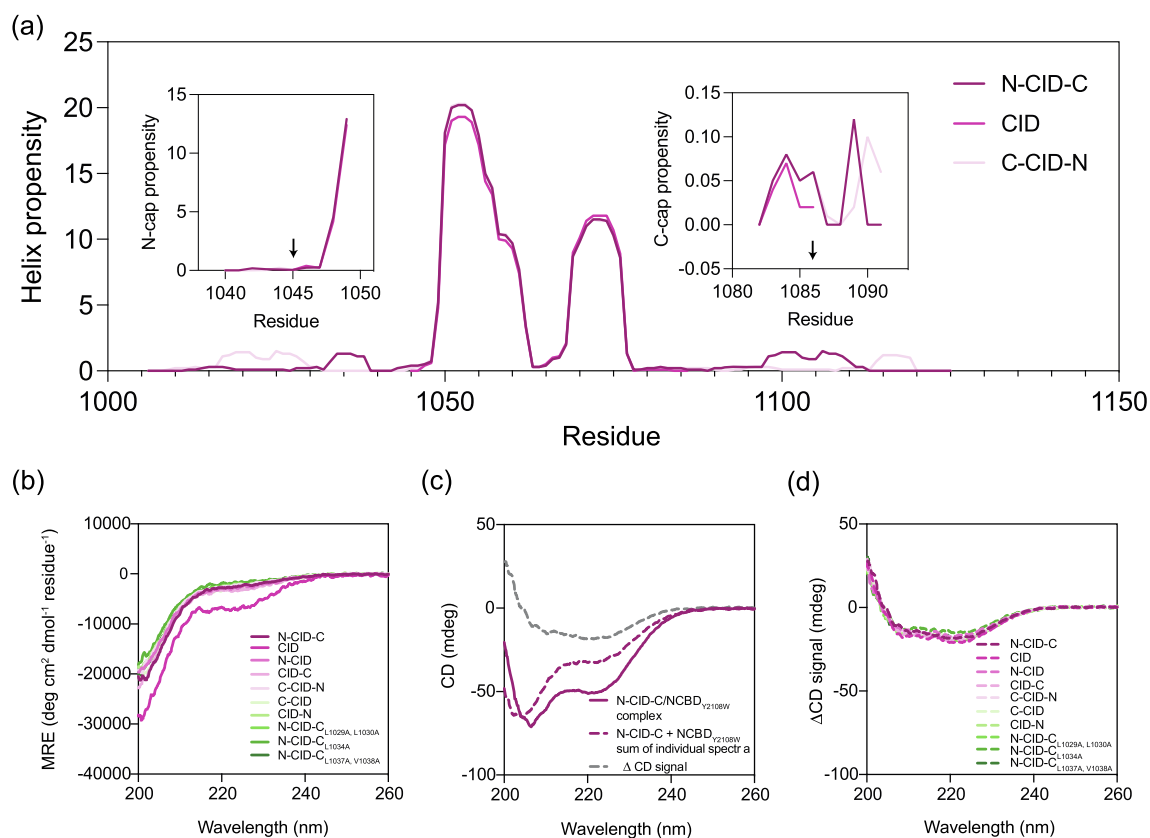
### Circular dichroism shows that the flanking regions are disordered in the complex with NCBD

First, we calculated the helix propensity on a residue level for three CID constructs using the AGADIR algorithm.<sup>31</sup> Indeed, CID has some tendency for helix formation, most notably in the  $\alpha_1$

helix. However, neither the N-terminal nor the C-terminal region displays any significant propensity to form a helix (Figure 2(a)). Next, we recorded circular dichroism (CD) spectra of all CID variants in the free state (Figure 2(b)). Spectra are shown as mean residue ellipticity (MRE), which allows comparison of the overall helicity of the constructs. All CID variants showed a characteristic CD spectrum with a prominent minimum at 200 nm and a weak signal at 222 nm, typical for IDPs. Compared to the longer constructs, the short CID had a slightly larger signal at 222 nm, indicating more helical structure on average. In agreement with the AGADIR prediction, the flanking regions apparently have little inherent residual structure and therefore reduce the MRE at 222 nm for the whole construct. To estimate the increase in helicity upon complex formation we recorded CD spectra of 20  $\mu$ M NCBD<sub>Y2108W</sub> (see next section for explanation of Y2108W) mixed with an equal concentration of the respective CID variant. We then compared the complex spectrum to the sum of the individual CD spectra of 20  $\mu$ M NCBD<sub>Y2108W</sub> and 20  $\mu$ M CID variant, as exemplified for N-CID-C (Figure 2(c)). We found a substantial increase in helicity upon complex formation, indicated by an increase in the CD signal at 222 nm and a slight shift of the peak at 205 nm to 208 nm compared to the sum of the individual spectra. The difference spectrum clearly shows the random coil to  $\alpha$ -helix transition in the complex, corresponding to the three  $\alpha$ -helices that are formed by CID in the complex. The difference spectra for CID variants with flanking regions were very similar to that of the short CID/NCBD complex and show an equal or slightly smaller extent of helix formation (Figure 2(d)). The comparable increase in helix formation upon binding and folding suggest, within error of the CD experiment, that the helical content of the binding region (residues 1045–1086) in all CID constructs in the free state is similar. Thus, experimental data (Figure 2(b)-(d)) and AGADIR predictions (Figure 2(a)) are most consistent with a scenario where intrinsic disorder remains in the flanking regions upon binding and that only residues within the CID region 1045–1086 increase the helix content upon binding with NCBD. However, we cannot experimentally rule out the less likely situation where the flanking regions induce helicity in the free state of CID (1045–1086) and then fold to compensate exactly for the observed overall change in helix content upon binding (Figure 2(d)).

### Stopped flow experiments show that flanking regions increase the affinity between CID and NCBD

Next, we measured the binding kinetics of the different CID variants to NCBD using stopped-flow fluorescence spectroscopy (Figure 3) as described in the Materials section. NCBD<sub>Y2108W</sub>, engineered with a Trp residue at position 2108, has been established previously in our lab as a variant



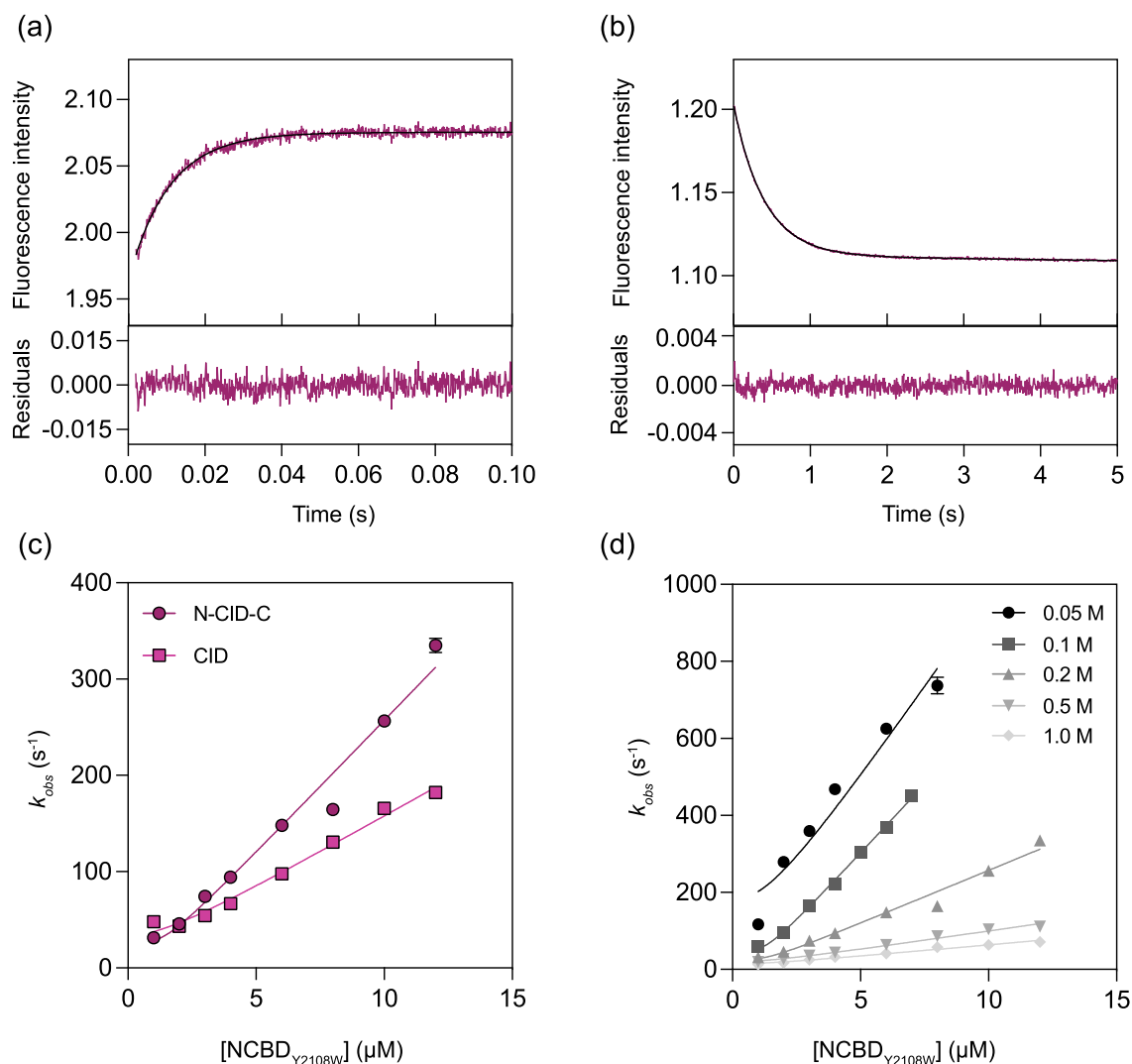
**Figure 2. Helical propensity and content of free and bound CID variants.** (a) AGADIR predictions suggest that the helical propensity in the flanking regions are low and that they do not contribute to increased N- or C-capping propensity. Consequently, the helical propensity of the binding region (CID, residues 1045–1086) is virtually identical for CID, N-CID-C and C-CID-N. The arrows indicate the start and stop, respectively, of the CID 1045–1086 variant. (b) Spectra of free CID variants. Note that the spectra are shown in MRE instead of the raw CD signal to compare the average helicity per amino acid. The shortest construct comprising the binding region (CID, residues 1045–1086) displays the highest MRE showing that the flanking regions have lower helical content in agreement with the AGADIR predictions in panel (a). (c) Helix formation upon binding for N-CID-C and NCBD<sub>Y2108W</sub>. Protein complex (solid dark purple), the sum of the CD spectra for free NCBD<sub>Y2108W</sub> and N-CID-C (dashed dark purple) and the difference between the two spectra (dashed grey), which shows the amount of helix formed upon binding. (d) Difference spectra between free and bound CID for all variants. The difference spectra are highly similar, indicating that the amount of helix formed upon binding is approximately the same for all CID variants. All spectra were recorded in 20 mM sodium phosphate (pH 7.4) and 150 mM NaCl at 4 °C.

suitable for kinetic studies, where the fluorescence yield of the tryptophan is used as a reporter for binding.<sup>22</sup>

A single exponential equation was fitted to all kinetic transients, yielding the observed rate constant  $k_{\text{obs}}$  (Figure 3(a)). To obtain  $k_{\text{on}}$  values,  $k_{\text{obs}}$  values were plotted against the concentration of NCBD<sub>Y2108W</sub> and the general equation for association of two molecules<sup>32</sup> (Eq. (1)) was fitted to the data (Figure 3(c)-(d)). By varying NCBD<sub>Y2108W</sub> and keeping the CID variant at a constant concentration, any errors in the determination of CID concentration will not affect the association rate constant  $k_{\text{on}}$ . The error in NCBD<sub>Y2108W</sub> concentration is low because of the Trp, minimizing the error in  $k_{\text{on}}$ . Furthermore, the same NCBD<sub>Y2108W</sub> stock was used for experiments on the different

CID variants making the comparison highly accurate, since any error in the NCBD<sub>Y2108W</sub> concentration will cancel out in the comparison between CID variants. Dissociation rate constants were recorded in separate displacement experiments (Figure 3(b)) (see Materials and Methods). The apparent two state means we can use the determined rate constants to calculate the affinity of the respective CID/NCBD<sub>Y2108W</sub> complex as  $K_D = k_{\text{off}}/k_{\text{on}}$  (Table 1, Supplementary excel file).

We performed kinetic experiments for seven CID variants with different arrangements of flanking regions: the short CID, the longest N-CID-C, two variants with truncated C- or N-termini, N-CID and CID-C, and three swapped variants C-CID-N, CID-N and C-CID (Figure 1). To assess a hydrophobic region in the N-terminal flanking



**Figure 3. Stopped flow kinetic experiments used to determine rate constants and affinity.** (a) Binding of 1  $\mu\text{M}$  N-CID-C to 4  $\mu\text{M}$  NCBD<sub>Y2108W</sub>. The kinetic trace was fitted to a single exponential function. (b) Displacement of the N-CID-C/NCBD<sub>Y2108W</sub> (each 1  $\mu\text{M}$ ) with 30  $\mu\text{M}$  NCBD<sub>WT</sub>. The transient was fitted to a single exponential function with a slope. The term describing the slope was added to account for a temperature-related artifact on the stopped-flow instrument. (c) Comparison of the recorded  $k_{\text{obs}}$  values of the N-CID-C and CID variant. (d) Observed rate constants of N-CID-C/NCBD<sub>Y2108W</sub> complex formation at different ionic strengths as indicated in the figure. Eq. (1) was fitted to experimental  $k_{\text{obs}}$  values in (c) and (d).

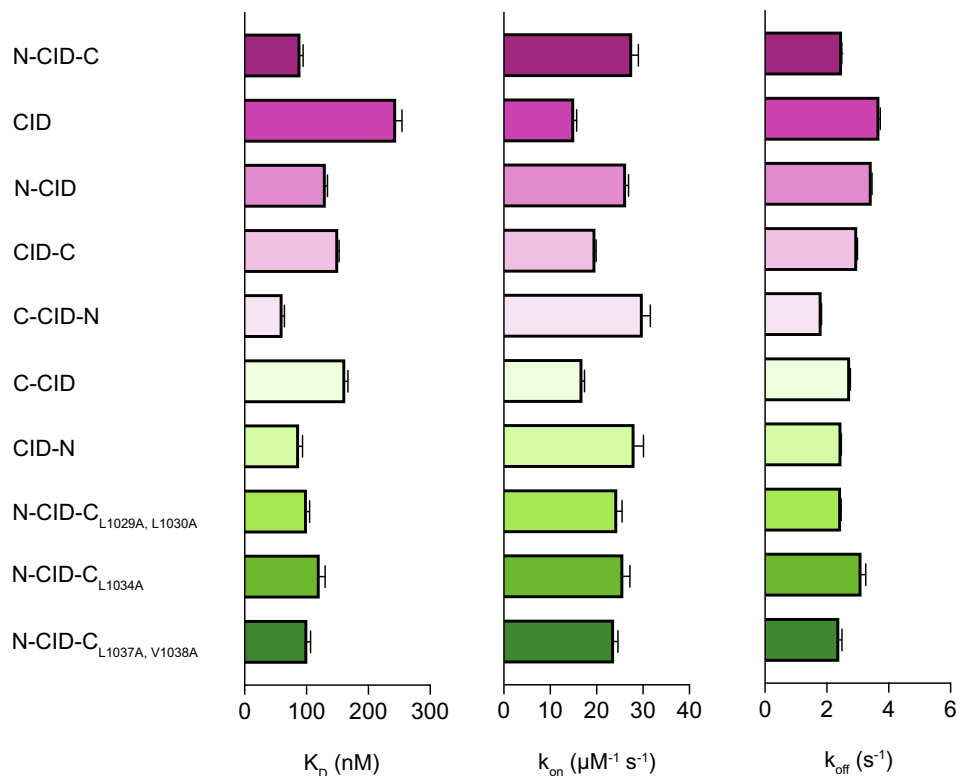
region we included three site-directed mutational variants of N-CID-C: N-CID-C<sub>L1029A L1030A</sub>, N-CID-C<sub>L1034A</sub>, and N-CID-C<sub>L1037A V1038A</sub>. Moreover, we subjected each of these ten variants to experiments at different ionic strength to investigate the role of electrostatic interactions involving the flanking regions. Overall, we found that the short CID displayed both the lowest association and the highest dissociation rate constant of all tested flanking region CID variants resulting in a 3-fold lower affinity as compared to the long N-CID-C (Figure 4, Table 1, Supplementary excel file). The N-CID and CID-C variants were intermediate in affinity. The C-CID-N

variant, with swapped N- and C-termini displayed a slightly higher affinity for NCBD<sub>Y2108W</sub> as compared to N-CID-C, demonstrating that the structural context of the flanking regions can influence the affinity, but that the interactions appear non-specific. In agreement with these results, both CID-N and C-CID displayed slightly higher or similar affinity (depending on ionic strength) as compared with N-CID and CID-C, respectively. There is limited cross-talk between the N- and C-terminus expressed as coupling free energy,  $\Delta\Delta\Delta G$ , obtained from a comparison between CID, N-CID-C, N-CID and CID-C (−0.02 to 0.18 kcal mol<sup>−1</sup>, Supplementary excel file).



**Table 1 Rate constants and affinity of the interaction between CID variants and NCBD<sub>Y2108W</sub>.** Data were obtained at an ionic strength of 0.2 M.  $K_D$  values were calculated from  $k_{off}/k_{on}$  with a propagated fitting error. The small error in  $k_{off}$  results from high accuracy and precision in displacement experiments, which are not affected by errors in concentration. Parameters obtained under all ionic strength conditions are presented in the supplementary excel file.

Construct	$k_{on}$ ( $\mu\text{M}^{-1}\text{s}^{-1}$ ) $\pm$ SEM	$k_{off}$ ( $\text{s}^{-1}$ ) $\pm$ SEM	$K_D$ (nM) $\pm$ SEM
N-CID-C	28 $\pm$ 1.4	2.5 $\pm$ 0.01	90 $\pm$ 5
CID	15 $\pm$ 0.6	3.7 $\pm$ 0.04	240 $\pm$ 10
N-CID	26 $\pm$ 0.6	3.4 $\pm$ 0.02	130 $\pm$ 3
CID-C	20 $\pm$ 0.2	3.0 $\pm$ 0.02	150 $\pm$ 2
C-CID-N	30 $\pm$ 1.7	1.8 $\pm$ 0.01	61 $\pm$ 4
C-CID	17 $\pm$ 0.5	2.7 $\pm$ 0.02	160 $\pm$ 5
CID-N	28 $\pm$ 2.0	2.5 $\pm$ 0.01	88 $\pm$ 6
N-CID-C <sub>L1029A, L1030A</sub>	24 $\pm$ 1.1	2.5 $\pm$ 0.01	100 $\pm$ 5
N-CID-C <sub>L1034A</sub>	26 $\pm$ 1.5	3.1 $\pm$ 0.15	120 $\pm$ 9
N-CID-C <sub>L1037A, V1038A</sub>	24 $\pm$ 0.9	2.4 $\pm$ 0.10	100 $\pm$ 6

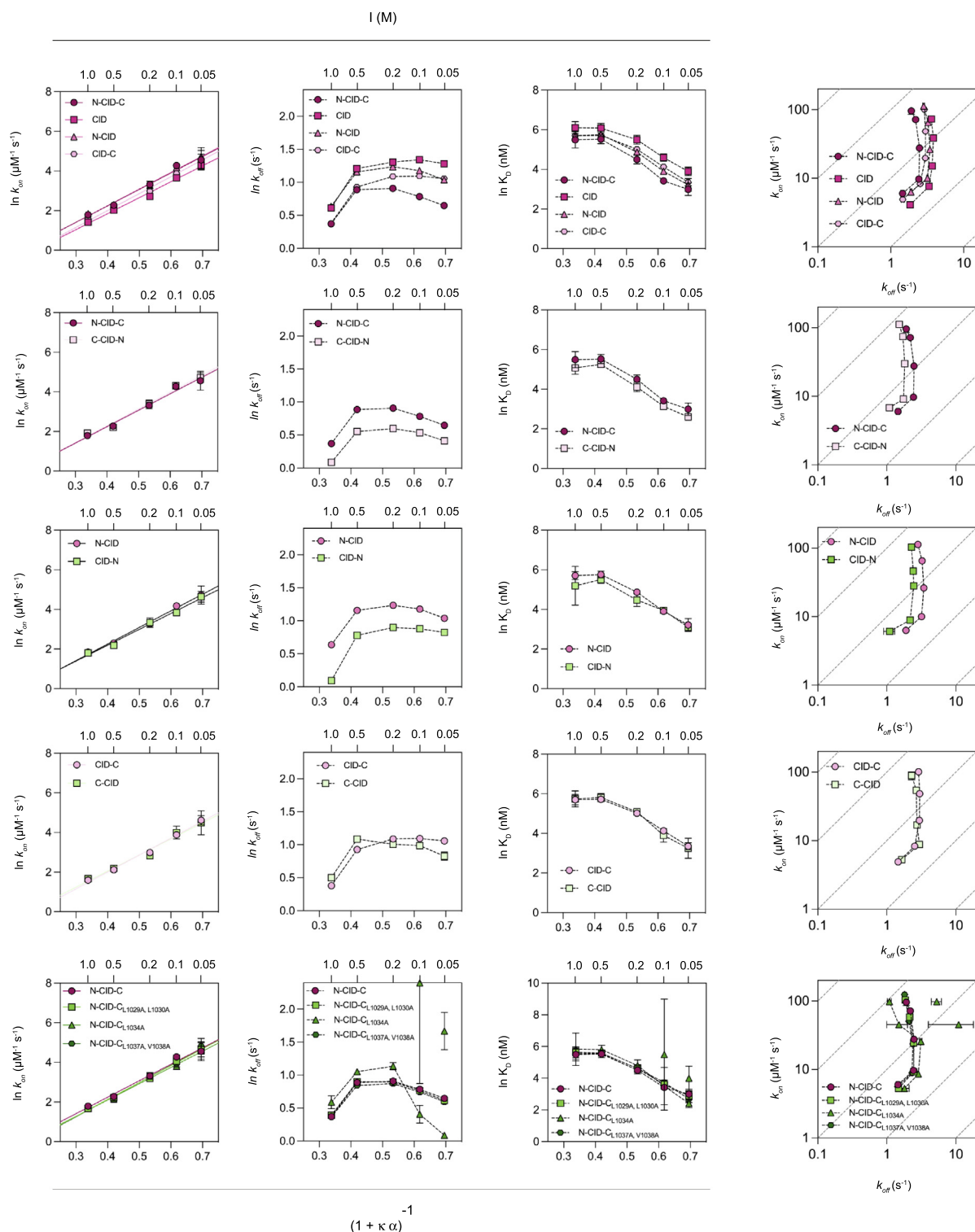


**Figure 4. Rate and dissociation constants from stopped flow experiments.** Dissociation ( $K_D$ ), association rate constant ( $k_{on}$ ) and dissociation rate constant ( $k_{off}$ ) for the different NCOA3 CID constructs. The parameters were obtained using stopped-flow experiments in 20 mM sodium phosphate, 150 mM NaCl and at 4 °C. All parameters are given with errors in the supplementary excel file.

### Charge interactions do not contribute to the increase in affinity by flanking regions

Electrostatic interactions are known to be the dominant factor to accelerate protein–protein association rates beyond the diffusion limit.<sup>33</sup> The association of CID and NCBD is driven by electrostatic interactions as shown by its strong dependence on ionic strength,<sup>21,34</sup> although the final complex is characterized by a large hydrophobic interface.<sup>19</sup> To investigate the role of electrostatics

in the effect caused by the flanking sequences, we recorded association and dissociation kinetics as a function of ionic strength using NaCl ( $I$  = 50 mM, 100 mM, 200 mM, 500 mM and 1.0 M, respectively) (Figure 5, Figure S2, Supplementary excel file). At sufficiently high ionic strength, ions are able to screen charged residues, thereby shielding attractive or repulsive electrostatic interactions between molecules.<sup>35</sup> Over the range from 50 mM to 1.0 M the association rate con-



**Figure 5. Dependence of rate and equilibrium constants on ionic strength.** The constants were subjected to Debye Hückel analysis, which estimates the basic association rate constant in the absence of ionic interactions (at high ionic strength). Data were obtained using stopped-flow fluorimetry at 4 °C in 20 mM sodium phosphate pH 7.4. The ionic strength varied between 0.050–1.0 M using NaCl. Error bars represent standard errors from the fitting or propagated standard errors. For  $k_{\text{off}}$ , the error bars denote the standard deviation of two separate  $k_{\text{obs}}$  values from displacement experiments. All parameters are given in the supplementary excel file.

stant of the short CID and NCBD<sub>Y2108W</sub> decreased 18-fold from 73  $\mu\text{M}^{-1} \text{s}^{-1}$  to 4.1  $\mu\text{M}^{-1} \text{s}^{-1}$ , while the dissociation rate constant only decreased 2-fold from 3.6  $\text{s}^{-1}$  to 1.8  $\text{s}^{-1}$ . Other CID variants displayed a very similar behavior, including a large ionic strength dependency for  $k_{\text{on}}$  (Figure S2, Supplementary excel file). We performed a Debye Hückel analysis to gain more insights into the interaction of the different CID variants and the origin of the increased association rate constant provided by the flanking regions. According to Eq. (2) a linear relationship between  $\ln k_{\text{on}}$  and  $(1 + \kappa\alpha)^{-1}$  yields the basal rate constant as the intercept at  $(1 + \kappa\alpha)^{-1} = 0$  and the electrostatic potential  $U$  as the slope of the curve (Figure 5). We determined the basal rate constant of CID as 0.26  $\mu\text{M}^{-1} \text{s}^{-1}$  and the electrostatic potential as 18.5  $\text{kJ mol}^{-1}$ . The large electrostatic potential reflects the strong influence of ionic charges on the association rate constant. Both the basal rate constant and the electrostatic potential were very similar for all CID constructs (Table 2, supplementary excel file). Indeed, most of the differences in the association rate constants between the short CID and any other variant were consistent over the whole range of ionic strength. Overall, the dissociation rate constant was much less affected by ionic strength, but displays more variation among the CID variants. This is best illustrated by the comparison of native N-CID-C and swap variants. For example, the increased affinity of the swapped C-CID-N variant compared to the N-CID-C version is maintained over all ionic strength conditions and solely caused by a reduced dissociation rate constant (Figure 5). For most CID constructs, the dissociation rate constant increased slightly between 0.05 and 0.2 M ionic strength. The dissociation rate constant reached a plateau or decreased slightly for most variants between 0.2 and 0.5 M ionic strength, whereas an abrupt decrease in the dissociation rate constant was observed for all variants above 0.5 M. The observed increase in the dissociation rate con-

stant up to 0.5 M ionic strength could be explained by screening of charges on NCBD and CID (see Discussion). The abrupt decrease in the dissociation rate constant above 0.5 M ionic strength may be ascribed to a stabilizing effect on the hydrophobic interactions between NCBD and CID. Alternatively, it might indicate a change in the rate-limiting step of dissociation. We have previously observed an additional kinetic phase (20–40  $\text{s}^{-1}$ ) for the interaction between NCBD and CID at  $\approx 1$  M ionic strength.<sup>22,36</sup> We did not observe this kinetic intermediate in our present experiments at high ionic strength, possibly due to a high total fluorescence resulting from NCBD<sub>Y2108W</sub> in excess. Nevertheless, accumulation of an alternative bound state at high salt concentrations might explain the decrease in the observed  $k_{\text{off}}$  at high ionic strength.

We also note that the salt dependence for  $k_{\text{off}}$  appears slightly different for the different CID variants. We observe a small but consistent difference in the salt dependence for N-CID-C as compared to CID-C, and for N-CID as compared to CID (Figure S5). This indicates that residues in the N-terminal disordered flanking regions have a small but detectable effect on stabilizing the bound state. In contrast to this, the salt dependence of  $k_{\text{off}}$  is highly similar for N-CID-C and N-CID, and for CID and CID-C, respectively, showing that the C-terminal flanking region does not contribute favorable electrostatic interactions in the bound complex.

In conclusion, while the interaction of CID and NCBD overall is characterized by strong electrostatic contributions, the flanking regions have little impact on the electrostatic rate enhancement. Hence, the flanking regions do not in general modulate the binding reaction via electrostatic interactions although a small effect on the dissociation rate constant was observed.

### Hydrophobic interactions in the flanking regions may contribute to the affinity

To address the role of the potential leucine-rich motif in the N-terminal flanking region we designed three variants with substitutions at hydrophobic residues, L1029A/L1030A, L1034A, and L1037A/V1038A, and subjected them to kinetic binding experiments. Intriguingly, the variants displayed slightly lower affinity towards NCBD<sub>Y2108W</sub> except at the lowest ionic strength (50 mM) where the affinity increased for all three variants. If we assume additivity, as observed for the N- and C-terminal linker variants, the combined effect of the five mutations would give a 1.5–1.8 fold lower affinity at higher ionic strengths ( $I = 100 \text{ mM}$ – $1.0 \text{ M}$ ) and a 2.7-fold higher affinity at  $I = 50 \text{ mM}$ . We may speculate that the mutations have two effects: substitution with Ala promotes helical propensity and thus affinity, but gives weaker hydrophobic interactions. Any positive effect of the Ala mutations on affinity would be

**Table 2 Basal rate constants and electrostatic potentials for the interaction between CID variants and NCBD<sub>Y2108W</sub>.** Basal rate constants ( $k_{\text{on, basal}}$ ) and electrostatic potentials ( $U$ ) were determined from data presented in Figure 5. We present  $U$  with three significant numbers to highlight the very small differences between CID constructs.

Construct	$k_{\text{on, basal}}$ ( $\mu\text{M}^{-1} \text{s}^{-1}$ )	$U$ ( $\text{kJ mol}^{-1}$ )
N-CID-C	$0.34 \pm 0.10$	$19.1 \pm 1.4$
CID	$0.25 \pm 0.05$	$18.6 \pm 1.1$
N-CID	$0.33 \pm 0.07$	$19.3 \pm 1.0$
CID-C	$0.24 \pm 0.04$	$19.7 \pm 1.0$
C-CID-N	$0.33 \pm 0.11$	$19.5 \pm 1.6$
C-CID	$0.30 \pm 0.10$	$18.8 \pm 1.7$
CID-N	$0.36 \pm 0.09$	$18.5 \pm 1.1$
N-CID-C <sub>L1029A, L1030A</sub>	$0.28 \pm 0.03$	$19.6 \pm 0.7$
N-CID-C <sub>L1034A</sub>	$0.30 \pm 0.05$	$18.9 \pm 0.8$
N-CID-C <sub>L1037A, V1038A</sub>	$0.26 \pm 0.05$	$20.0 \pm 1.2$



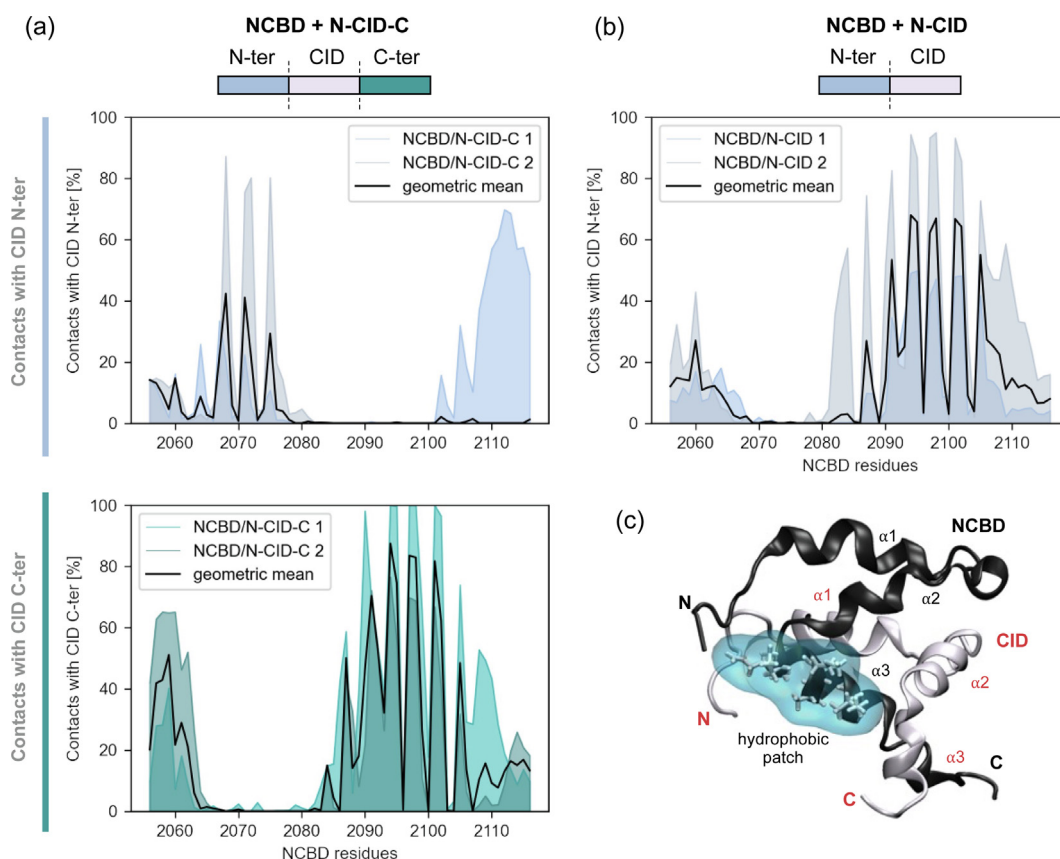
offset by loss of hydrophobic interactions made by the Leu residues that are promoted at high ionic strength.

### MD simulations suggest that a hydrophobic patch in NCBD is responsible for the increased affinity provided by the flanking regions

To shed further light on the molecular basis regulating the effect of the flanking regions we performed molecular dynamics (MD) simulations. We ran two couples of plain MD simulations: two simulations (replicates) of NCBD in complex with N-CID-C and two replicates of NCBD with N-CID. The starting structures did not contain any contacts between NCBD and the N- or C-terminus of CID to avoid bias towards specific contacts. Simulations were evolved for 1 to 2  $\mu$ s and analyzed to identify possible interacting regions between NCBD and CID terminal tails.

Firstly, we focused on which NCBD residues can form contacts with CID flanking regions. For each NCBD residue we computed the percentage of frames in which it was in contact with at least one

residue of either the CID N-terminus or the CID C-terminus (Figure 6). This contact probability was computed individually for each simulation, then the geometric mean was calculated to highlight only the common contacts: contacts present in both the replicates were considered as contributing to binding, while contacts appearing in only one of the two simulations were considered as random. The NCBD/N-CID-C simulations revealed that the CID N-terminus engages mainly random contacts with NCBD, while the CID C-terminus contacts a well-defined group of NCBD residues. These include residues P2094, Q2095, M2097, A2098 and I2101, forming a hydrophobic patch on NCBD (Figure 6(c)). Interestingly, in NCBD/N-CID simulations (Figure 6(b)), where the CID C-terminal tail is missing, the contacts with the NCBD hydrophobic patch are stably engaged by the N-terminus. This result suggests that the increased binding affinity observed for the N-CID-C construct is likely due to contacts engaged by the flanking regions with the NCBD hydrophobic patch. According to our simulations, when both

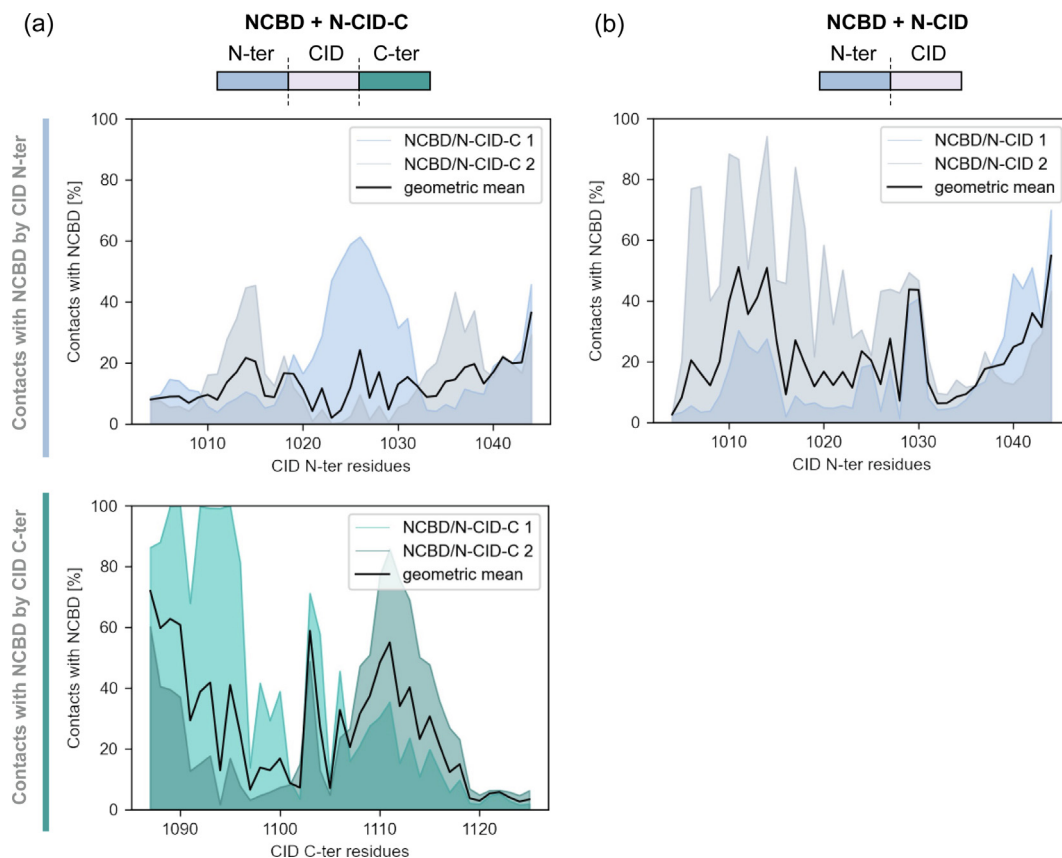


**Figure 6. NCBD residues contacting CID flanking regions.** (a) For the two NCBD/N-CID-C simulations, the probability of each NCBD residue to be in contact with CID N-terminus (top panel) or CID C-terminus (bottom panel) is reported. (b) For the two NCBD/N-CID simulations, the probability of each NCBD residue to be in contact with CID N-terminus is reported. The black line represents the geometric mean of the two probabilities. (c) Representative structure of the NCBD/CID complex, in which the hydrophobic patch on NCBD helix  $\alpha_3$  is highlighted with licorice and a light-blue surface.

the N- and C-terminal tails are present, the hydrophobic patch is preferentially contacted by the CID C-terminus, which is spatially closer (Figure 6(c)). Nevertheless, if the C-terminal tail is missing (N-CID), the hydrophobic patch can make stable interactions with the CID N-terminus.

Interestingly, the binding affinities measured for the different CID variants to NCBD indicate that the CID-N displays a binding affinity comparable to N-CID-C and higher than both N-CID, C-CID and CID-C at high ionic strength. This suggests that the leucine-rich motif in the N-terminus could ideally engage contacts with the NCBD hydrophobic patch; still the residues in the C-terminal flanking regions are favored in contacting the patch due to their spatial proximity. To verify whether we could identify an analogous binding preference on the side of CID flanking regions, we performed a similar analysis focusing on the percentage of contacts engaged by each residue from the CID N- or C-terminus with NCBD (Figure 7). In the NCBD/N-CID-C simulations (Figure 7(a)), in line with the observation that the NCBD hydrophobic patch is stably contacted only

by the C-terminal tail, we identified a preference only for a few CID C-terminal residues: these include the first residues of the C-terminal tail (i.e., 1087–1090, ALEP, which indeed are in close proximity to the NCBD hydrophobic patch) plus M1103, L1110 and Y1111. Vice versa, in the NCBD/N-CID simulations (Figure 7(b)), where the CID N-terminal flanking region contacts the NCBD hydrophobic patch, we identified CID N-terminal residues, which are preferentially involved in these inter-molecular contacts: W1010, P1011, G1013, M1014, L1029, L1030 and L1044. Out of these, L1029 and L1030 are part of the  $\Phi\Phi\text{xx}\Phi$  motif. In this context, it is worth noting that the binding preference of the CID flanking regions displays a less clear signal with respect to the one shown by NCBD (compare Figure 6 and Figure 7), suggesting that any hydrophobic motif on the CID flanking regions could contribute to improving the affinity by contacting the well-defined NCBD hydrophobic patch. Overall, the MD simulations support the idea of transient or fuzzy interactions between CID flanking regions and NCBD.



**Figure 7. CID flanking region residues contacting NCBD.** (a) For the two NCBD/N-CID-C simulations, the probability of each CID N-terminal residue (top panel) or CID C-terminal residue (bottom panel) to be in contact with NCBD is reported. (b) For the two NCBD/N-CID simulations, the probability of each CID N-terminal residue to be in contact with NCBD is reported. The black line represents the geometric mean of the two probabilities.

## Probing the role of the hydrophobic patch on NCB in the interactions with CID flanking regions

To assess the MD simulations experimentally we made NCB variants in which three amino acid residues in the hydrophobic patch were changed to Ala, NCB<sub>Q2095A M2097A I2101A</sub>. We hypothesized that the difference between short and long CID should decrease upon binding to this triple mutant variant. However, the triple mutant yielded stopped-flow traces with low signal-to-noise and low kinetic amplitudes, likely due to a destabilization of the protein complex and/or a change in the environment around the fluorescent Trp probe on NCB. Therefore, we turned to isothermal titration calorimetry (ITC) for measuring the affinities. While ITC works well in this affinity range (1–10  $\mu$ M) and is not dependent on a fluorescent probe, the accuracy and precision is lower than in the kinetic experiments, and small differences may be difficult to discern. We performed experiments with NCB<sub>Q2095A M2097A I2101A</sub> and the CID variants N-CID-C and short CID (Figure S3). The ITC data showed a 2–3-fold change in affinity between N-CID-C and CID for binding to NCB<sub>Q2095A M2097A I2101A</sub>. This change in affinity is similar to the difference observed for N-CID-C and CID in the kinetic experiments ( $\approx$  3-fold). Thus, these data do not support a role for the residues in the hydrophobic patch in the interactions with flanking regions of CID. However, like the wild-type residues, the Ala residues in the triple mutant could provide a hydrophobic patch for the flanking regions to interact with, although smaller. Therefore, we also tested an NCB variant in which the same three residues had been substituted with polar Ser residues. This NCB<sub>Q2095S M2097S I2101S</sub> variant was highly precipitation-prone and exhibited a complex binding isotherm in ITC experiments with N-CID-C and CID, indicating formation of a higher order complex and thus precluding a robust interpretation of the results. In conclusion, the triple Ala mutant displayed a destabilized interaction, resulting in high observed rate constants and experimental noise in the kinetic experiments, and did not permit a quantitative assessment of the role of the interaction between CID flanking regions and the hydrophobic patch in NCB.

## Discussion

In cellular signaling networks, most proteins interact via an intrinsically disordered short linear binding motif in one protein and an interaction domain in the other one. The binding motif usually folds into a well-defined ordered structure upon binding to the interaction domain. However, the interaction motif is often located in a longer disordered region and the flanking regions of the

motif do not fold upon binding. These flexible flanking regions could potentially influence the affinity via mechanisms that have only recently been investigated in detail and are not fully understood.<sup>10–11,37–39</sup> Our CD experiments confirm the presence of disorder in the flanking regions of free as well as bound CID, suggesting that the higher affinity provided by the flanking regions is neither caused by preformed structure nor folding upon binding. Furthermore, the kinetic data show that the effect is not based on long-range electrostatic interactions. This finding was somewhat surprising, considering that NCB is highly positively charged and the disordered flanking regions of CID are weakly negatively charged at neutral pH based on the amino acid sequence. However, previous measurements of the zeta potential of NCB showed that its actual charge is only  $\sim +0.5$  at an ionic strength of 0.05 M (KCl). Furthermore, when increasing the KCl concentration, NCB obtains a net negative charge around 0.2 M ionic strength.<sup>40</sup> This charge reversal is likely due to formation of a shell of chloride ions around the highly positive NCB and might explain the overall low contribution of electrostatic interactions from the disordered flanking regions, and the observed small contribution to  $k_{off}$  at low ionic strength. Thus, the strong dependence on electrostatics observed for the NCB/CID association kinetics appears to be mediated by specific interactions between residues in the minimal interaction domains,<sup>22</sup> rather than non-specific electrostatic interactions.

Instead, our data suggest that the flanking regions form non-specific hydrophobic and/or polar interactions with NCB. These interactions are context dependent, but local sequence is not crucial as shown by the termini-swapped CID variants. However, the N-terminal and the C-terminal flanking regions have a similar overall amino acid composition, suggesting that the higher affinity may be mediated via amino acid-specific, but structurally non-constrained fuzzy interactions, as has been hypothesized for other IDP systems.<sup>41</sup> The presence of such transient contacts is hard to detect experimentally, but can be indirectly deduced by modulation of the affinity of a complex as shown here for the interaction between NCB and CID. The increase in affinity provided by the flanking regions likely results from transient interactions in both the N- and C-terminal regions.

On a more general level, what are the functional implications of the observed effects from the disordered flanking regions on affinity? Is this a phenomenon that arises due to non-specific interactions in essentially all protein–protein interactions or a specific interaction under selection? More data are required to address these questions, but we can speculate that the presence of a hydrophobic hotspot on the NCB surface, mediating the transient contacts with the

disordered flanking regions, suggests that the interactions might be functionally relevant. Fuzzy interactions in the flanking regions may in this way fine tune the affinity of two proteins, regardless of how optimized the binding interface is. Since interaction motifs are often found in larger disordered regions our findings provide support for the idea of structural plasticity and fuzzy interactions in protein complexes as a general mechanism by which intrinsically disordered regions modulate the affinity for their binding partners. Due to lower structural constraints in the flanking regions as compared to the binding interface, there is likely more room for sequence variation from mutation on which natural selection can act to fine tune the affinity.

## Material and methods

### Expression and purification

A modified pRSET vector containing an N-terminal hexa-histidine-tagged lipolyt fusion tag, followed by a thrombin cleavage site (LVPRGS) and either a variant of CID or NCBD (wtNCBD, NCBD<sub>Y2108W</sub>) were purchased from GenScript Biotech. The plasmid was transformed into *Escherichia coli* BL21(DE3) pLysS cells. Cells were grown at 37 °C in 1 × TY media to an OD<sub>600</sub> of 0.6 and expression was induced by addition of 1 mM Isopropyl β-D-1- thiogalactopyranoside. Protein expression was carried out at 18 °C overnight, after which the cells were harvested and resuspended in lysis buffer (30 mM Tris/HCl, pH 7.9, 150 mM NaCl). Cells were sonicated on ice and after centrifugation for 60 min at 20,000 g and 4 °C the supernatant was filtered (0.22 μm filter; Sarstedt) and loaded on a Ni-NTA column, equilibrated with washing buffer (30 mM Tris/HCl, pH 7.9, 500 mM NaCl). The column was washed extensively with buffer and the His-tagged fusion protein was eluted with 250 mM imidazol. The eluate was dialyzed against 30 mM Tris/HCl (pH 7.9); 200 mM NaCl over night. Subsequently, the fusion protein was cleaved with 60 μL of thrombin (GE Healthcare) at room temperature overnight. To remove His-tagged lipolyt, uncleaved fusion protein and thrombin, the mixture was first filtered and then passed over a Ni-NTA column and a Benzamidine column. The flow through was collected and the pH adjusted to acidic conditions with HCl. The protein was then loaded onto a C-8 or RESOURCE RPC reversed phase chromatography column (from Grace Davison Discovery Sciences and Cytiva, respectively) and eluted with a gradient from 100 % H<sub>2</sub>O and 0.1 % TFA to 70 % acetonitrile; 30 % H<sub>2</sub>O; 0.1 % TFA. The presence of protein was monitored by measuring the absorption at 215 nm and 280 nm. After elution, fractions containing protein were diluted with ddH<sub>2</sub>O, frozen on dry ice, lyophilized and stored at −20 °C. Purity of the protein was

checked by SDS-PAGE and identity confirmed by MALDI-TOF.

### Circular dichroism spectroscopy

Far-UV circular dichroism (CD) spectra were recorded on a Jasco J-1500CD spectropolarimeter. Protein samples were prepared at concentrations of 20 μM in 20 mM sodium phosphate (pH 7.4) and 150 mM NaCl and placed in a quartz cuvette (Hellma, Müllheim, Germany) with a path length of 1 mm. For each protein sample 5 spectra were recorded, averaged and a separate buffer spectrum was subtracted. All spectra were recorded at 4 °C.

### Stopped flow spectroscopy

Stopped-flow experiments were performed on an upgraded SX-17 MV stopped-flow spectrometer (Applied Photophysics). Samples were prepared at 2 times the intended final concentration and placed in the two syringes. Thus, NCBD<sub>Y2108W</sub> was mixed rapidly with the respective CID variant and the change in fluorescence upon binding was monitored over time. The engineered tryptophan of NCBD was excited at 280 nm and the emission was monitored over time with a 320 nm long pass filter. Binding experiments were performed at 4 °C with a constant concentration of the CID variant of 1 μM, while NCBD<sub>Y2108W</sub> was varied over 1–12 μM. For a typical measurement, i.e., at each NCBD<sub>Y2108W</sub> concentration, 10–20 individual traces were averaged into one trace. In the experimental range of concentrations used, all binding kinetics were in good agreement with a simple two-state mechanism, i.e., they were well described by a single exponential function. Previous studies reported an additional slow phase observed in association experiments with fixed concentrations of NCBD and increasing excess concentrations of CID<sub>1018-1088</sub>.<sup>22</sup> We did not observe this kinetic phase in experiments with excess NCBD<sub>Y2108W</sub>. However, kinetic phases may not be observable under different experimental conditions, either because they occur on a different time scale, or with a marginal amplitude. Hence, a single exponential equation was fitted to the data to obtain a  $k_{obs}$  value for the respective concentration pair. Eq. (1) was then fitted to the data.<sup>23</sup>

$$k_{obs} = \sqrt{k_{on}^2([A]_0 - [B]_0)^2 + 2k_{on}k_{off}([A]_0 + [B]_0) + k_{off}^2} \quad (1)$$

This yields the association rate constant  $k_{on}$  and dissociation rate constant  $k_{off}$ .

For displacement experiments 1 μM of the CID variant and 1 μM of NCBD<sub>Y2108W</sub> were mixed and preincubated for 30 min at 4 °C prior to the experiment. The complex was placed in one syringe and rapidly mixed with a large excess of wild-type NCBD (with a Tyr at position 2108). The dissociation of the complex was monitored by



measuring the decrease of Trp fluorescence over time as wild-type NCBD replaces NCBD<sub>Y2108W</sub> in the complex. The large excess of NCBD<sub>WT</sub> prevents the reformation of the fluorescent complex and at sufficiently high concentrations of non-fluorescent competitor the  $k_{\text{obs}}$  values reach a plateau. Here, the observed rate constant reflects only the dissociation of the initial fluorescent complex and fitting a single exponential then yields the dissociation rate constant  $k_{\text{off}}$ . Dissociation rate constants  $k_{\text{off}}$  were calculated as the average of two  $k_{\text{obs}}$  values at high wild-type NCBD concentration in displacement experiments. The standard deviation of these two experiments is given as the error. Since  $k_{\text{off}}$  is a first-order rate constant, which is not concentration dependent, displacement experiments provide highly accurate estimates of  $k_{\text{off}}$ .

In displacement experiments with the variants C-CID and CID a slower second phase with an amplitude of opposite sign was observed and a double exponential was used for fitting to account for this phase. The reported  $k_{\text{obs}}$  values correspond to the first phase. At ionic strength 0.05 and 0.10 M, the N-CID-C<sub>L1034A</sub> variant displayed double exponential kinetics where the two phases had amplitudes with the same sign. Since it was not clear in this case which phase reflected the dissociation, both  $k_{\text{off}}$  values were plotted and used to calculate two separate  $K_D$  values (Figure 5). Furthermore, a term describing a slope was in some cases added to the exponential function to account for a temperature-related drift of the signal in the stopped-flow instrument not related to the dissociation kinetics.

### Debye Hückel analysis

Kinetics were recorded at different ionic strengths of 0.050 M, 0.10 M, 0.20 M, 0.50 M and 1.0 M, respectively. The ionic strength was adjusted with sodium chloride. A Debye Hückel like approximation (Eq. (2)) was used to estimate the basal rate constant  $k_{\text{on,basal}}$  and the electrostatic potential  $U$  for the interaction.<sup>35</sup>

$$\ln k_{\text{on}} = \ln k_{\text{on,basal}} - \frac{U}{RT} \times \frac{1}{1 + \kappa\alpha} \quad (2)$$

where  $R$  is the universal gas constant,  $T$  the temperature,  $\kappa$  is the inverse Debye Hückel length and  $\alpha$  the minimal distance of approach. For analysis of the ionic strength dependency  $k_{\text{on}}$  values were plotted against  $(1 + \kappa\alpha)^{-1}$ , with  $\alpha$  being approximated as 6 Å.<sup>24–25</sup>

### Molecular dynamics simulations

Two couples of plain Molecular Dynamics (MD) simulations were performed using GROMACS 2021<sup>42</sup> and PLUMED 2<sup>43</sup>: two simulations (replicates) of NCBD in complex with N-CID-C and two replicates of NCBD with N-CID. Two different start-

ing structures of the NCBD/N-CID-C complex were generated from the PDB file 1kbh, adding the missing terminal residues with Modeller<sup>44</sup> and introducing the NCBD Y2108W mutation. The same structures were used for NCBD/N-CID simulations removing the C-terminal residues. To avoid bias toward specific contacts of the CID flanking regions with NCBD residues, the starting CID N- and C-terminal tails did not display any contacts with NCBD. The system was described using the Amber99SB-disp force-field with the TIP4PD water model.<sup>45</sup> The initial models were solvated with water molecules in a dodecahedron box initially 0.6 nm larger than the protein in each direction (note that in the starting conformation the tails are completely extended). The system was neutralized with a salt concentration of 100 mM NaCl. After an initial energy minimization to a maximum force of 100 kJ/mol/nm, the solute was equilibrated under NVT condition at the temperature of 310 K for 50 ps using the Bussi thermostat<sup>46</sup>, then stochastic cell rescaling barostat<sup>47</sup> was used to equilibrate the system in the NPT ensemble to the target pressure of 1 atm for 200 ps, maintaining the temperature at 310 K. Electrostatics was treated by using the particle mesh Ewald scheme<sup>48</sup> with a short-range cutoff of 0.9 nm and a Fourier grid spacing of 0.12 nm; van der Waals interaction cutoff was set to 0.9 nm. LINCS algorithm<sup>49</sup> was used to constraint bonds involving hydrogen atoms using a simulation time step of 2 fs and a multiple time step algorithm for long-range interactions. The equilibration phase was followed by production runs, 1 to 2  $\mu$ s long, in the NPT ensemble, maintaining temperature and pressure at the values of 310 K and 1 atm respectively, using the Bussi thermostat and the stochastic cell rescaling barostat. For the analysis, the first 100 ns were discarded; then only the frames containing at least one contact, defined by two non-hydrogen atoms being closer than 0.55 nm, between NCBD and CID N- or C-terminus were considered (that was the case for at least 60% of frames in all the simulations).

### Acknowledgements

This work was supported by the Swedish Research Council (2020-04395 to P.J.), and partially supported by Fondazione Cariplo (CoronAid to C.P. and C.C.), Fondazione Telethon (GGP19134 to C.C.), and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2033 (390677874 to C.H.). C.P. and C.C. acknowledge PRACE for awarding them access to Piz Daint at CSCS, Switzerland.



## Author contributions

EK and PJ conceived the project. EK, JS, and PJ designed the wet experiments and analyzed data. EK, JS and EA performed experiments. CP and CC designed and performed the computational experiments and analyzed data. EK, JS, CP, CC and PJ wrote the paper. All authors contributed to revision of the first draft and interpretation of data. PJ supervised the project.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmb.2022.167643>.

Received 8 April 2022;

Accepted 16 May 2022;

Available online 20 May 2022

### Keywords:

intrinsically disordered proteins;  
flanking regions;  
protein interactions;  
affinity;  
binding motif

† Equal contribution to the experimental work.

## References

- Bondos, S.E., Dunker, A.K., Uversky, V.N., (2022). Intrinsically disordered proteins play diverse roles in cell signaling. *Cell Commun. Signal* **20**, 20.
- Staby, L., O'Shea, C., Willemoës, M., Theisen, F., Kragelund, B.B., Skriver, K., (2017). Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochem. J.* **474**, 2509–2532.
- Fuxreiter, M., Simon, I., Friedrich, P., Tompa, P., (2004). Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* **338**, 1015–1026.
- Tompa, P., Davey, N.E., Gibson, T.J., Babu, M.M., (2014). A million peptide motifs for the molecular biologist. *Mol. Cell* **55**, 161–169.
- Gianni, S., Freiburger, M.I., Jemth, P., Ferreira, D.U., Wolynes, P.G., Fuxreiter, M., (2021). Fuzziness and frustration in the energy landscape of protein folding, function, and assembly. *Acc. Chem. Res.* **54**, 1251–1259.
- Tompa, P., Fuxreiter, M., (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* **33**, 2–8.
- Gianni, S., Dogan, J., Jemth, P., (2016). Coupled binding and folding of intrinsically disordered proteins: what can we learn from kinetics? *Curr. Opin. Struct. Biol.* **36**, 18–24.
- Naudi-Fabra, S., Tengo, M., Jensen, M.R., Blackledge, M., Milles, S., (2021). Quantitative description of intrinsically disordered proteins using single-molecule FRET, NMR, and SAXS. *J. Am. Chem. Soc.* **143**, 20109–20121.
- Bugge, K., Brakti, I., Fernandes, C.B., Dreier, J.E., Lundsgaard, J.E., Olsen, J.G., Skriver, K., Kragelund, B. B., (2020). Interactions by disorder - a matter of context. *Front. Mol. Biosci.* **7**, 110.
- Staby, L., Due, A.D., Kunze, M.B.A., Jørgensen, M.L.M., Skriver, K., Kragelund, B.B., (2021). Flanking disorder of the folded  $\alpha\alpha$ -hub domain from radical induced cell death1 affects transcription factor binding by ensemble redistribution. *J. Mol. Biol.* **433**, 167320.
- Theisen, F.F., Staby, L., Tidemand, F.G., O'Shea, C., Prestel, A., Willemoës, M., Kragelund, B.B., Skriver, K., (2021). Quantification of conformational entropy unravels effect of disordered flanking region in coupled folding and binding. *J. Am. Chem. Soc.* **143**, 14540–14550.
- Chong, S.-H., Ham, S., (2019). Folding free energy landscape of ordered and intrinsically disordered proteins. *Sci. Rep.* **9**, 14927.
- Chen, H., Lin, R.J., Schiltz, R.L., Chakravarti, D., Nash, A., Nagy, L., Privalsky, M.L., Nakatani, Y., Evans, R.M., (1997). Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. *Cell* **90**, 569–580.
- Kjaergaard, M., Teilum, K., Poulsen, F.M., (2010). Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. *PNAS* **107**, 12535–12540.
- Kjaergaard, M., Andersen, L., Nielsen, L.D., Teilum, K., (2013). A folded excited state of ligand-free nuclear coactivator binding domain (NCBD) underlies plasticity in ligand recognition. *Biochemistry* **52**, 1686–1693.
- Dogan, J., Toto, A., Andersson, E., Gianni, S., Jemth, P., (2016). Activation barrier-limited folding and conformational sampling of a dynamic protein domain. *Biochemistry* **55**, 5289–5295.
- Zosel, F., Mercadante, D., Nettek, D., Schuler, B., (2018). A proline switch explains kinetic heterogeneity in a coupled folding and binding reaction. *Nat. Comm.* **9**, 3332.
- lešmantavičius, V., Dogan, J., Jemth, P., Teilum, K., Kjaergaard, M., (2014). Helical propensity in an intrinsically disordered protein accelerates ligand binding. *Angew. Chem. Int. Ed.* **53**, 1548–1551.
- Demarest, S.J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H.J., Evans, R.M., Wright, P.E., (2002). Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **415**, 549–553.
- Demarest, S.J., Deechongkit, S., Dyson, H.J., Evans, R. M., Wright, P.E., (2004). Packing, specificity, and mutability at the binding interface between the p160 coactivator and CREB-binding protein. *Protein Sci.* **13**, 203–210.
- Sturzenegger, F., Zosel, F., Holmstrom, E.D., Buholzer, K. J., Makarov, D.E., Nettek, D., Schuler, B., (2018). Transition path times of coupled folding and binding reveal the formation of an encounter complex. *Nat. Comm.* **9**, 4708.
- Dogan, J., Schmidt, T., Mu, X., Engström, Å., Jemth, P., (2012). Fast association and slow transitions in the

- interaction between two intrinsically disordered protein domains. *J. Biol. Chem.* **287**, 34316–34324.
23. Karlsson, E., Paissoni, C., Erkelens, A.M., Tehranizadeh, Z.A., Sorgenfrei, F.A., Andersson, E., Ye, W., Camilloni, C., Jemth, P., (2020). Mapping the transition state for a binding reaction between ancient intrinsically disordered proteins. *J. Biol. Chem.* **295**, 17698–17712.
  24. Dogan, J., Mu, X., Engström, Å., Jemth, P., (2013). The transition state structure for coupled binding and folding of disordered protein domains. *Sci. Rep.* **3**, 2076.
  25. Karlsson, E., Andersson, E., Dogan, J., Gianni, S., Jemth, P., Camilloni, C., (2019). A structurally heterogeneous transition state underlies coupled binding and folding of disordered proteins. *J. Biol. Chem.* **294**, 1230–1239.
  26. Ganguly, D., Zhang, W., Chen, J., (2012). Synergistic folding of two intrinsically disordered proteins: searching for conformational selection. *Mol. Biosyst.* **8**, 198–209.
  27. Knott, M., Best, R.B., (2014). Discriminating binding mechanisms of an intrinsically disordered protein via a multi-state coarse-grained model. *J. Chem. Phys.* **140**, 175102.
  28. Liu, X., Chen, J., Chen, J., (2019). Residual structure accelerates binding of intrinsically disordered ACTR by promoting efficient folding upon encounter. *J. Mol. Biol.* **431**, 422–432.
  29. Toto, A., Camilloni, C., Giri, R., Brunori, M., Vendruscolo, M., Gianni, S., A, T., C, C., R, G., M, B., M, V., S, G., (2016). Molecular recognition by templated folding of an intrinsically disordered protein. *Sci. Rep.* **6**, 21994.
  30. Hultqvist, G., Åberg, E., Camilloni, C., Sundell, G.N., Andersson, E., Dogan, J., Chi, C.N., Vendruscolo, M., Jemth, P., (2017). Emergence and evolution of an interaction between intrinsically disordered proteins. *Elife* **6** e16059.
  31. Muñoz, V., Serrano, L., (1997). Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* **41**, 495–509.
  32. Malatesta, F., (2005). The study of bimolecular reactions under non-pseudo-first order conditions. *Biophys. Chem.* **116**, 251–256.
  33. Vijayakumar, M., Wong, K.Y., Schreiber, G., Fersht, A.R., Szabo, A., Zhou, H.X., (1998). Electrostatic enhancement of diffusion-controlled protein-protein association: comparison of theory and experiment on barnase and barstar. *J. Mol. Biol.* **278**, 1015–1024.
  34. Dogan, J., Jonasson, J., Andersson, E., Jemth, P., (2015). Binding rate constants reveal distinct features of disordered protein domains. *Biochemistry* **54**, 4741–4750.
  35. Schreiber, G., Haran, G., Zhou, H.-X., (2009). Fundamental aspects of protein-protein association kinetics. *Chem. Rev.* **109**, 839–860.
  36. Karlsson, E., Andersson, E., Jones, N.C., Hoffmann, S.V., Jemth, P., Kjaergaard, M., (2019). Coupled binding and helix formation monitored by synchrotron-radiation circular dichroism. *Biophys. J.* **117**, 729–742.
  37. Troilo, F., Bonetti, D., Bignon, C., Longhi, S., Gianni, S., (2019). Understanding intramolecular crosstalk in an intrinsically disordered protein. *ACS Chem. Biol.* **14**, 337–341.
  38. Miskei, M., Horvath, A., Vendruscolo, M., Fuxreiter, M., (2020). Sequence-based prediction of fuzzy protein interactions. *J. Mol. Biol.* **432**, 2289–2303.
  39. Schrag, L.G., Liu, X., Thevarajan, I., Prakash, O., Zolkiewski, M., Chen, J., (2021). Cancer-associated mutations perturb the disordered ensemble and interactions of the intrinsically disordered p53 transactivation domain. *J. Mol. Biol.* **433**, 167048.
  40. Japrun, D., Dogan, J., Freedman, K.J., Nadzeyka, A., Bauerdick, S., Albrecht, T., Kim, M.J., Jemth, P., Edel, J.B., (2013). Single-molecule studies of intrinsically disordered proteins using solid-state nanopores. *Anal. Chem.* **85**, 2449–2456.
  41. Fuxreiter, M., (2019). Fold or not to fold upon binding - does it really matter? *Curr. Opin. Struct. Biol.* **54**, 19–25.
  42. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., Lindahl, E., (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25.
  43. Tribello, G.A., Bonomi, M., Branduardi, D., Camilloni, C., Bussi, G., (2014). PLUMED 2: New feathers for an old bird. *Comp. Phys. Comm.* **185**, 604–613.
  44. Webb, B., Sali, A., (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics* **54**.
  45. Robustelli, P., Piana, S., Shaw, D.E., (2018). Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.*, 115.
  46. Bussi, G., Donadio, D., Parrinello, M., (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101.
  47. Bernetti, M., Bussi, G., (2020). Pressure control using stochastic cell rescaling. *J. Chem. Phys.* **153**, 114107.
  48. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G., (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593.
  49. Hess, B., (2008). P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122.