BIOCHEMISTRY

Structure and dynamics conspire in the evolution of affinity between intrinsically disordered proteins

Per Jemth[1]*, Elin Karlsson[1], Beat Vögeli[2], Brenda Guzovsky[3], Eva Andersson[1], Greta Hultqvist[4], Jakob Dogan[5], Peter Güntert[6,7,8], Roland Riek[6], Celestine N. Chi[1]*

In every established species, protein-protein interactions have evolved such that they are fit for purpose. However, the molecular details of the evolution of new protein-protein interactions are poorly understood. We have used nuclear magnetic resonance spectroscopy to investigate the changes in structure and dynamics during the evolution of a protein-protein interaction involving the intrinsically disordered CREBBP (CREB-binding protein) interaction domain (CID) and nuclear coactivator binding domain (NCBD) from the transcriptional coregulators NCOA (nuclear receptor coactivator) and CREBBP/p300, respectively. The most ancient low-affinity "Cambrian-like" [540 to 600 million years (Ma) ago] CID/NCBD complex contained less secondary structure and was more dynamic than the complexes from an evolutionarily younger "Ordovician-Silurian" fish ancestor (ca. 440 Ma ago) and extant human. The most ancient Cambrian-like CID/NCBD complex lacked one helix and several interdomain interactions, resulting in a larger solvent-accessible surface area. Furthermore, the most ancient complex had a high degree of millisecond-to-microsecond dynamics distributed along the entire sequences of both CID and NCBD. These motions were reduced in the Ordovician-Silurian CID/NCBD complex and further redistributed in the extant human CID/NCBD complex. Isothermal calorimetry experiments show that complex formation is enthalpically favorable and that affinity is modulated by a largely unfavorable entropic contribution to binding. Our data demonstrate how changes in structure and motion conspire to shape affinity during the evolution of a protein-protein complex and provide direct evidence for the role of structural, dynamic, and frustrational plasticity in the evolution of interactions between intrinsically disordered proteins.

INTRODUCTION

The web of protein-protein interactions that governs cell function is very complex, often involving competing interactions as well as spatial and temporal regulation of the proteins. Since most of this regulation has evolved over tens or hundreds of millions of years, it is likely that any new protein-protein interaction resulting from, for example, gene duplication, point mutation, horizontal gene transfer, or up-regulation of expression will interfere with the biochemistry of the cell in a non-beneficial way and disappear through purifying selection. However, sometimes, new interactions are sufficiently beneficial to be retained in the population by natural selection. This is what happened sometime between 600 and 540 million years (Ma) ago for two intrinsically disordered protein (IDP) domains: the nuclear coactivator binding domain (NCBD) from CREB-binding protein [CREBBP, also called CBP, and its paralog p300 (CREBBP/p300)] and the CREBBP interaction domain (CID) from the nuclear coactivator protein [nuclear receptor coactivator 1 (NCOA1), NCOA2, and NCOA3, also called Src1, Tif2, and ACTR, respectively]. The CREBBP/p300 and NCOA proteins contain histone acetyl transferase domains and are involved in transcriptional regulation, for example, by activating steroid receptors (1). The tighter

association between NCOA and CREBBP/p300 likely resulted in a more efficient transcription, which was sufficiently beneficial to be retained by natural selection in an ancestral population leading to present-day deuterostome animals.

Little is known about the molecular details of protein evolution as it happened in nature. Today, the large number of genome and transcriptome sequencing efforts of extant species allows reconstruction of ancient proteins from evolutionary ancestors, which provides the possibility of studying the putative historical sequence of events leading up to present-day proteins (2–5). We recently published the reconstruction, biophysical characterization, and molecular dynamics simulation of CID and NCBD domains from animal species living approximately 390 to 600 Ma ago (6) and found that the affinity between CID and NCBD was relatively low in an early deuterostome ancestor, which likely existed at some point 540 to 600 Ma ago (i.e., before or at the beginning of the Cambrian period). However, the affinity of the CID/NCBD protein-protein complex had probably increased, along with decreased conformational heterogeneity, to that of present-day CID/NCBD complexes already in an early ancestor of present-day fish living around or before 500 Ma ago during Cambrian and at the end of Ordovician/beginning of Silurian at the time of the two whole-genome duplications in the vertebrate lineage around 440 Ma ago. This affinity was maintained in the bony fish/tetrapod ancestor 390 Ma ago and is retained in present-day species (6).

To gain a deeper insight into the molecular details shaping the evolution of this protein-protein interaction, we have used here nuclear magnetic resonance (NMR) to obtain structures of three historical CID/NCBD complexes and analyzed the dynamics of the respective interaction by Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion NMR spectroscopy. We find that the most ancient Cambrian-like CID/NCBD complex lacks some of the structural features of the younger complexes and that there are a reduction and reshuffling of the

[1]Department of Medical Biochemistry and Microbiology, Uppsala University, BMC Box 582, SE-75123 Uppsala, Sweden. [2]Department of Biochemistry and Molecular Genetics, University of Colorado at Denver, 12801 East 17th Avenue, Aurora, CO 80045, USA. [3]Protein Physiology Lab, FCEyN-Universidad de Buenos Aires, IQUIBICEN/CONICET, Intendente Güiraldes 2160, Ciudad Universitaria, C1428EGA Buenos Aires, Argentina. [4]Department of Pharmaceutical Biosciences, Uppsala University, BMC Box 591, SE-75124 Uppsala, Sweden. [5]Department of Biochemistry and Biophysics, Stockholm University, SE-10691 Stockholm, Sweden. [6]Laboratory of Physical Chemistry, ETH Zürich, ETH-Hönggerberg, Zürich, Switzerland. [7]Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, Goethe University, 60438 Frankfurt am Main, Germany. [8]Graduate School of Science, Tokyo Metropolitan University, Tokyo 192-0397, Japan.
*Corresponding author. Email: chi.celestine@imbim.uu.se (C.N.C.); per.jemth@imbim.uu.se (P.J.)

backbone dynamics as well as local frustration going from the low-affinity ancestral to the younger Ordovician-Silurian/extant human complexes. Furthermore, isothermal titration calorimetry (ITC) experiments show opposing enthalpic and entropic contributions to the total affinity. The structural heterogeneity and plasticity observed in our evolutionary snapshots are consistent with an affinity sensitive to changes in local frustration distribution and thermodynamics and shed light on fundamental questions regarding the evolution of protein-protein interactions.

RESULTS

### Structures of CID/NCBD complexes through evolution

To investigate how structure and conformational dynamics modulate protein function in natural evolution, we solved three protein complexes of CID/NCBD domains spanning roughly 600 Ma of evolution (Protein Data Bank codes 6ES5, 6ES6, and 6ES7). The ancient NCBD and CID domain sequences were previously reconstructed, and the resurrected proteins were subjected to a biophysical analysis (6). Using our previous nomenclature, we denote these domains D/P NCBD, which was present in the common ancestor of present-day Eubilateria [i.e., deuterostome (D) and protostome (P) animals], a creature that likely lived at some point 540 to 600 Ma ago; 1R/2R NCBD and 1R CID, which were present in fish that lived close to the transition between Ordovician and Silurian ~440 Ma ago (1R and 2R are the first and second round of the whole-genome duplications that took place in the vertebrate lineage that later diversified into present-day vertebrates such as jawed fish, reptiles, and mammals) (7); and CREBBP NCBD and NCOA3 CID, which are two extant human domains (Fig. 1A). Since the CID domain could not be identified in species from protostome phyla (e.g., Arthropoda and Mollusca) and sequences were too few and different in the deuterostome phyla Echinodermata (e.g., sea stars and sea urchins) and Hemichordata (e.g., acorn worms), it was not possible to reconstruct a CID domain older than 1R and as ancient as the D/P NCBD. It is also important to point out that any reconstructed sequence is an approximation based on gene variants from selected present-day species for which sequence data are available. Nevertheless, on the basis of the data on alternative ancient variants, we deem our overall conclusion robust, namely, that the Cambrian-like complex has a significantly lower affinity than the two younger complexes around neutral pH (K_D, ~1 μM versus 100 nM; table S1). Thus, the three complexes analyzed in the present work consisted of the most likely CID and NCBD variants: (i) 1R CID with D/P NCBD (most ancient Cambrian-like complex), (ii) 1R CID with 1R/2R NCBD (the Ordovician-Silurian 1R/2R complex), and (iii) extant human CID/NCBD complex (from NCOA3 and CREBBP). Note that the two whole-genome duplications 1R and 2R occurred in a relatively short but undefined time span. Thus, while it is theoretically possible to reconstruct two CID variants (since three of four NCOA genes are present today), in practice, the 1R/2R should be treated as one evolutionary node. In addition, ITC binding data show that the affinities of 1R and 2R CID for 1R/2R NCBD are similar (6).

First, we determined the structure of the extant human CID/NCBD complex by NMR using nuclear Overhauser effect (NOE) distance restraints, measured ϕ dihedral angles restraints from scalar couplings, chemical shift–derived ϕ/ψ dihedral angles, and ^1D H-N residual dipolar couplings (RDCs) (8). The use of RDCs in structure calculations improves the overall quality by restraining the relative orientation of the structural elements (9). Comparison to a previously published NCOA3 CID/CREBBP NCBD structure (10) shows that the domains fold into

similar overall structures with comparable compactness and total solvent-accessible surface area (SASA) [5730 Å^2 (previous structure) versus 5940 Å^2 (our structure)]. A few minor differences between the structures likely result from the different lengths of the CID and NCBD constructs used in our structure compared to the previous one.

We then solved the structures of the most ancient Cambrian-like complex and the Ordovician-Silurian 1R/2R complex to understand how the CID/NCBD interaction has evolved over time. The overall topology of the resurrected ancient complexes was similar to that of the extant human one (Fig. 1, C to E). Furthermore, all three complexes had a rotational correlation time (τ_c) of approximately 5 ns, as expected for a complex of this size and indicating that the domains gain structure upon binding. This is markedly different from the free CID domains, which had rotational correlation times of less than 1 ns, consistent with a disordered structure (11). We also observe that the SASA decreased from the Cambrian-like (7280 Å^2) to the Ordovician-Silurian complex (6210 Å^2) and the extant human complex (5940 Å^2) (Fig. 2). This change in compactness is associated with changes in the orientation of the first and third α helices of NCBD (Nα1 and Nα3, respectively) and the third α helix of CID (Cα3) (Figs. 1 and 2 and fig. S1). 1R CID interacts with the Cambrian-like D/P NCBD using only Cα1 and Cα2, while, upon binding to Ordovician-Silurian 1R/2R NCBD, the third helix Cα3 makes more contacts with Nα2 of NCBD, permitting it to adopt a more helical structure (Figs. 1 to 3 and figs. S1 to S3). On more detailed inspection, we find that the additional contacts in this region in the extant human CID/NCBD complex result in an increase in helicity of both Nα3 and Cα3 (Figs. 2 and 3 and figs. S1 to S3). These additional structural elements increase the hydrophobic interaction area between CID and NCBD and provide new networks of interactions within the younger higher-affinity CID/NCBD complexes. For example, in the most ancient complex, Leu^{2074}, Lys^{2075}, Ser^{2076}, Leu^{2087}, and Ile^{2101} of Cambrian-like NCBD make direct contacts (as measured by NOEs) with Leu^{1049}, Ile^{1050}, and Ile^{1073} of 1R CID. In the Ordovician-Silurian 1R/2R complex, the number of detected direct NOEs tripled to 18, notably interactions between Ile^{2062}, Leu^{2067}, Gln^{2068}, Lys^{2072}, Gln^{2095}, Gln^{2103}, and Ala^{2105} of Ordovician-Silurian 1R/2R NCBD and Asp^{1044}, Glu^{1045}, Leu^{1048}, Leu^{1056}, Leu^{1067}, and Gln^{1080} of 1R CID. These interactions are maintained in the extant human complex except Asn^{1079} of NCOA3 CID taking the place for Gln^{1080} of 1R CID. These results are further supported by the change in the amide chemical shift ($\Delta\delta H^N$), which shows increased bending of the Nα3 and Cα3 helices as they displayed alteration of the H^N shift in a wave-like manner (fig. S4). Because bending will result in shorter hydrogen bonds between the carbonyl and amide groups of residues i, $i + 4$ at the concave side and longer hydrogen bonds at the convex side, amides at the concave side will show positive $\Delta\delta H^N$ while those at the convex side will show more negative values (12). This behavior results in repeats of the amide proton secondary chemical shifts with periodicity of three to four along the sequence.

Further interesting details are provided by the side chains of residues 1070 to 1080, which are partly solvent-exposed and remain partially unstructured in the most ancient Cambrian-like complex (Figs. 2A and 3 and fig. S1) but form part of the interaction surface in the Ordovician-Silurian 1R/2R complex (Fig. 3H). The interaction surface is further increased by interactions involving helix Nα3 of the two younger NCBD domains. Specifically, Ala^{2105} and Lys^{2107} of 1R/2R NCBD interact with Ile^{1073} and Asp^{1068} of 1R CID, respectively, and Thr^{2105} of human CREBBP NCBD forms an interaction with His^{1053} of NCOA3 CID. In the most ancient complex, residues in Nα3 make interactions with only part of Cα3 of 1R CID, while most of the Cα3
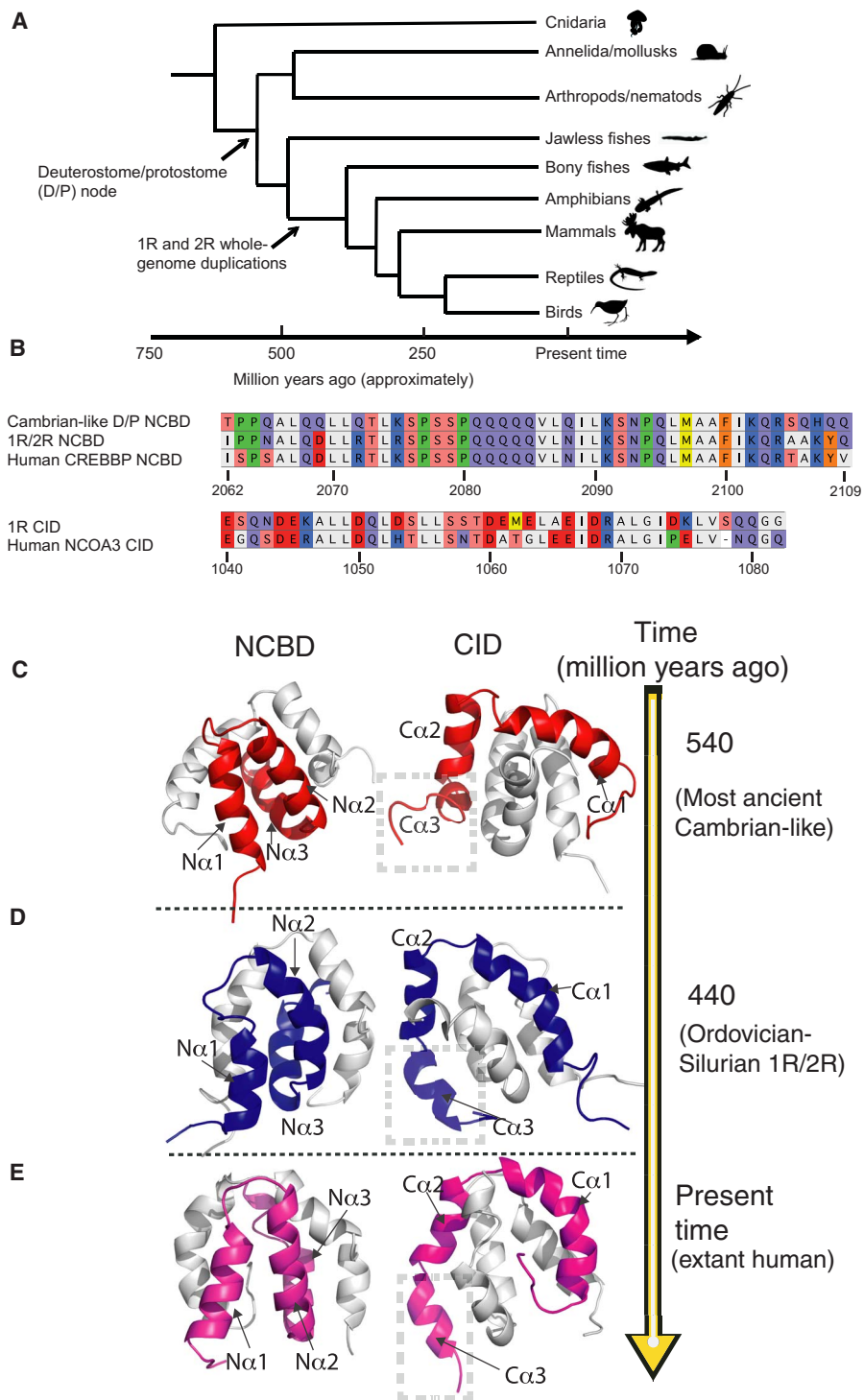
Fig. 1. Reconstructed ancestral sequences, phylogenetic tree ancestral, and extant CID/NCBD complexes. (**A**) The sequences were previously reconstructed (6) using 181 (NCBD) and 184 (CID) protein sequences from present-day protostome and deuterostome species of the animal phyla, superclasses, and classes depicted in the simplified phylogenetic tree. (**B**) The complexes used in this study are (i) 1R CID with D/P NCBD (most ancient Cambrian-like complex), (ii) 1R CID with 1R/2R NCBD (the Ordovician-Silurian 1R/2R complex), and (iii) extant human CID/NCBD complex (from NCOA3 and CREBBP, respectively). Note that the timing of the whole-genome duplications 1R and 2R is not resolved and might predate the divergence of jawed and jawless vertebrates (25). Animals were downloaded from phylopic.org. (**C**) The most ancient Cambrian-like complex (red) contains D/P NCBD and 1R CID. (**D**) The Ordovician-Silurian 1R/2R complex (blue) contains 1R/2R NCBD and 1R CID. (**E**) The extant human complex (magenta) contains human CREBBP NCBD and NCOA3 CID. In the three left panels, the solved structures of NCBD domains are displayed in cartoon representation and colored, while the CID domains are cartoon and gray. In the three right panels, the solved structures of CID domains are displayed as colored cartoons, and the NCBD domains are gray cartoons. The complexes in the left and right panels are rotated 180° with respect to each other.
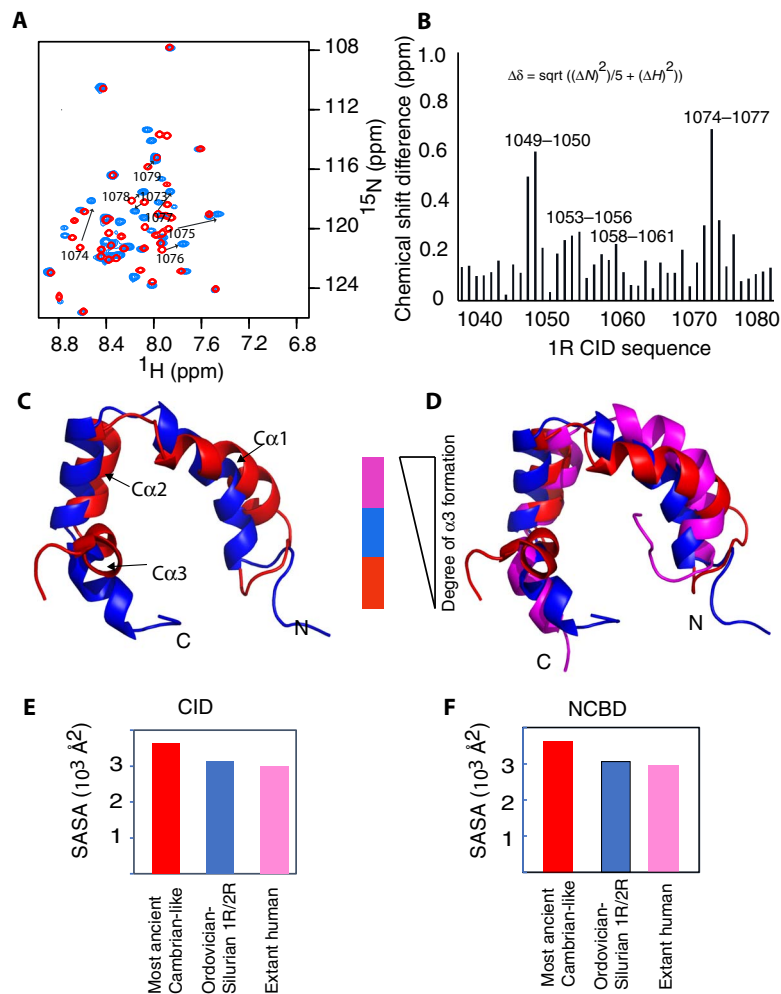
Fig. 2. Evolutionary snapshots highlighting structural heterogeneity. (**A**) Overlay of the heteronuclear single-quantum correlation spectra for 1R CID bound to Cambrian-like NCBD (red) and Ordovician-Silurian NCBD (blue). Arrows indicate the peak shifts of the amino acid residues 1073 to 1079 corresponding to the CID Cα3 helix (for clarity, only the assignments of residues 1073 to 1079 are shown; complete assignments can be found in fig. S7). In the most ancient low-affinity complex with Cambrian-like NCBD, these residues remain unstructured. However, in the younger high-affinity Ordovician-Silurian 1R/2R complex, the increased number of interactions in the region leads to increased α-helical content of Cα3 (see Fig. 4). The increase in dispersion for residues 1074 to 1077 is accompanied by an increase in the number of inter- and intramolecular NOEs (i, $i + 3$, indicative of helices; figs. S3 and S4), indicating that this region becomes more structured. ppm, parts per million. (**B**) Root mean square deviations (RMSDs) between the combined H^N/N chemical shifts for the complexes of 1R CID with Cambrian-like D/P NCBD and Ordovician-Silurian 1R/2R NCBD, respectively. Large structural changes are seen for residues Leu[1049] and Asp[1050] located in Cα1 and for residues Asp[1074], Lys[1075], Leu[1076], and Val[1077] in Cα3. sqrt, square root. (**C**) Superposition of structures of 1R CID bound to Cambrian-like NCBD (red) and Ordovician-Silurian NCBD (blue), respectively. (**D**) 1R CID bound to Cambrian-like NCBD (red) and Ordovician-Silurian NCBD (blue), and human NCOA3 CID bound to CREBBP NCBD (magenta). For clarity, only the CID domains are displayed in (C) and (D). (**E** and **F**) Plots of total SASA calculated from the respective complex and plotted separately for each domain, CID (E) and NCBD (F).

remains exposed to the solvent. Thus, the three historical complexes demonstrate how an extensive reorganization of structure and interactions, rather than specific side-chain substitutions in an intact structure, can lead to a higher affinity in a protein-protein complex.

Redistribution of motions accompanies the evolution of higher affinity in the CID/NCBD complex

Since dynamics on different time scales is an inherent property of proteins and a modulator of protein function, we next examined the evolution of dynamics within the three complexes using NMR relaxation dispersion CPMG experiments. These experiments can probe slow motions (i.e., in the microsecond-to-millisecond time scale) that are often relevant for function (*13*). Equilibrium interconversion between different protein conformations in this time window will

frequently result in an exchange contribution to the transverse relaxation rate constant (R_{ex}), causing additional line broadening of the respective NMR spectral resonances with respect to the fast motion–driven R_0. The exchange contributions, R_{ex}, depend on the relative populations of the exchanging states, the chemical shift difference ($\Delta\omega$) between them, and the rate constant of exchange (k_{ex}) (Eq. 1). We observed that the Cambrian-like complex exhibited a varying degree of slow motions over the entire protein sequences (Fig. 4). In the Ordovician-Silurian 1R/2R complex, slow motions were markedly reduced in CID with a few exceptions. In particular, helices Cα1, Cα2, and Cα3 contribute little or no slow motions. In the extant human complex, slow motions appear in other locations, resulting in conformational dynamics mostly in the loop region between Cα1 and Cα2 and in the entire Cα3 region, which experiences increased R_{ex} values. The chemical shift–derived order
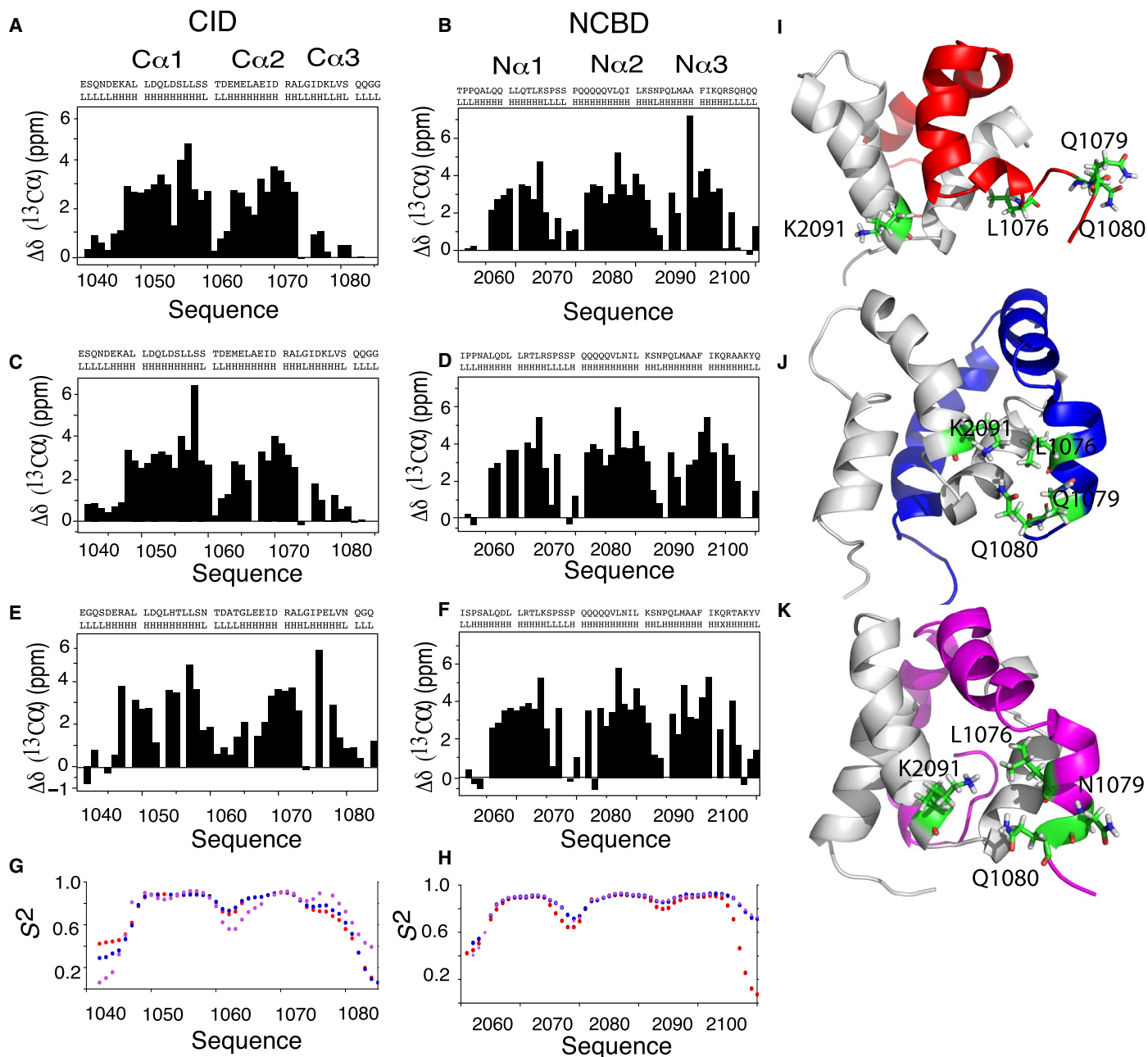
Fig. 3. Evolution of Cα3 and Nα3 helical content and conformational rearrangements within the Cα3 helix of CID domains during evolution. Plots of the difference of experimental Cα shifts from those of random coil values as a function of the amino acid sequences and TALOS prediction for ancestral and extant CID domains (**A**, **C**, and **E**) and NCBD domains (**B**, **D**, and **F**). The height of the bars indicates the degree of α helix formed with zero, indicating no α-helical content. There is a general increase in helical content for Cα3 and Nα3 as we evolve from Cambrian-like and Ordovician-Silurian (1R/2R) complexes to the present-day human CID/NCBD complex. (**G** and **H**) Chemical shift–based S^2 values show that the flexibility in the Cα3 and Nα3 helices is higher in the most ancient low-affinity Cambrian-like complex (red) than in the younger high-affinity Ordovician-Silurian 1R/2R (blue) and present-day human (magenta) complexes. Specific interactions and rearrangements of K2091 of Nα2 and L1076, Q1079, and Q1080 of Cα3 (side chains are numbered and colored green) for (**I**) Cambrian-like D/P NCBD (gray) with 1R CID (red), (**J**) Ordovician-Silurian 1R/2R NCBD bound to 1R CID (blue), and (**K**) the extant human complex between CREBBP NCBD and NCOA3 CID (magenta). The three highlighted Cα3 residues are mainly solvent exposed in the most ancient Cambrian-like complex, while specific interactions are seen in both the Ordovician-Silurian 1R/2R and the extant human complex. For clarity, the structures have been slightly reoriented in relation to each other to show the specific interactions.

parameters, which are sensitive to motions faster than the inverse chemical shift difference in Hertz, (ca. 10 ms), also showed increased millisecond motions in this region (Fig. 3G). It is reasonable that such reshuffling of slow motions partly compensates for the loss of motion associated with the polypeptide going from disorder to order upon

binding. In the two ancient complexes, the helix is not or only partially formed such that the conformational interconversion takes place in the fast-motion regime (faster than nanoseconds; Fig. 3G and fig. S4) (14). This process has been slowed down in the extant human complex because of the relatively stable Cα3 resulting in larger R_{ex}
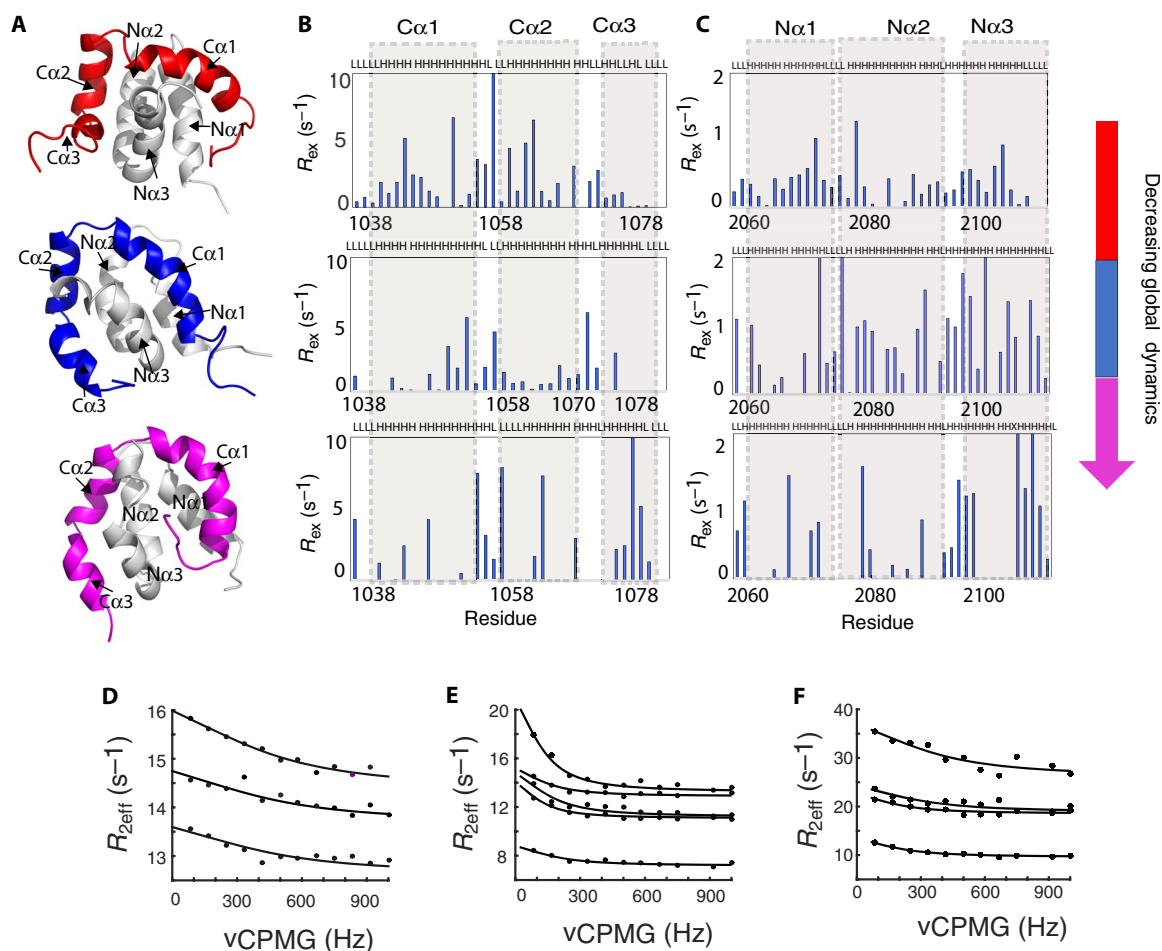
Fig. 4. Redistribution of slow motions and flexibility during the evolution of the CID/NCBD complex. Plots of R_{ex} from relaxation dispersion experiments as a function of position along the amino acid sequences for ancestral and extant CID domains (**B**) and NCBD domains (**C**). The CPMG probe (83 to 1000 Hz) used for these experiments detects conformational exchange occurring on the microsecond-to-millisecond time scale, and nonzero R_{ex} values therefore indicate the presence of motions in this time regime. Notably, all three complexes show a distinct distribution. (**A**) Cartoon representation of the respective complex of CID/NCBD domains displayed in a similar orientation. R_{2eff} determined as a function of different CPMG frequencies were fitted globally to Eq. 1 (see Methods), describing a two-site exchange process for selected residues. k_{ex} values of 3260 ± 1850 s^{-1}, 1800 ± 245 s^{-1}, and 2300 ± 1000 s^{-1} were obtained for (**D**) the most ancient Cambrian-like complex, (**E**) Ordovician-Silurian 1R/2R complex, and (**F**) the extant human CID/NCBD complex, respectively.

values (Fig. 4). The trends in dynamics are less clear for NCBD, but Nα1 and Nα2 display a loss in dynamics slower than nanoseconds when comparing the most ancient with the two younger complexes. R_{ex} in Nα3 changes from the N- to the C-terminal half of the helix. Thus, along with formation of more intermolecular CID/NCBD interactions as measured by NOEs, there is a general trend toward less and more localized conformational heterogeneity in the CID/NCBD complex upon going from the low-affinity Cambrian-like complex to the higher-affinity younger complexes (Figs. 3 and 4). We like to point out that localization and redistribution of motions could be highly dependent on particular side chains. Therefore, any erroneously predicted amino acid residue in the ancestral reconstruction could significantly influence the dynamics. Nevertheless, by comparing the Ordovician-Silurian 1R/2R complex with the extant human one, it is clear that high affinity is compatible with different dynamics profiles. Our data suggest that redistribution of motions may play a role in fine-tuning affinity upon mutations in diverging sequences. It is important to note that faster motions (nanoseconds to picoseconds) have also been implicated in function (*15, 16*). However, it was not

possible to perform model-free dynamic analysis of the free domains, which would be needed for a conversion of NMR relaxation data into thermodynamic parameters as the local motions cannot be easily separated from global tumbling. Nevertheless, we provide information on the backbone subnanosecond motions in terms of $R_2(R_{1\rho})/R_1$ (fig. S5). The results indicate that, mostly, the N-terminal, C-terminal, and loop regions with low $R_2(R_{1\rho})/R_1$ values exhibit motions in the nanosecond-to-picosecond time scale.

Thermodynamic basis for the different affinities of the CID/NCBD complexes

To investigate how the observed changes in structure, dynamics, and affinity are reflected in the binding thermodynamics, we performed ITC experiments for the three complexes at different pH and temperature (Fig. 5, fig. S5, and table S1). ITC experiments directly measure the change in enthalpy (ΔH) for the binding reaction and, from the shape of the titration profile, the equilibrium association constant (K_A). From these parameters, K_D ($=1/K_A$), ΔS, and ΔG can be calculated.
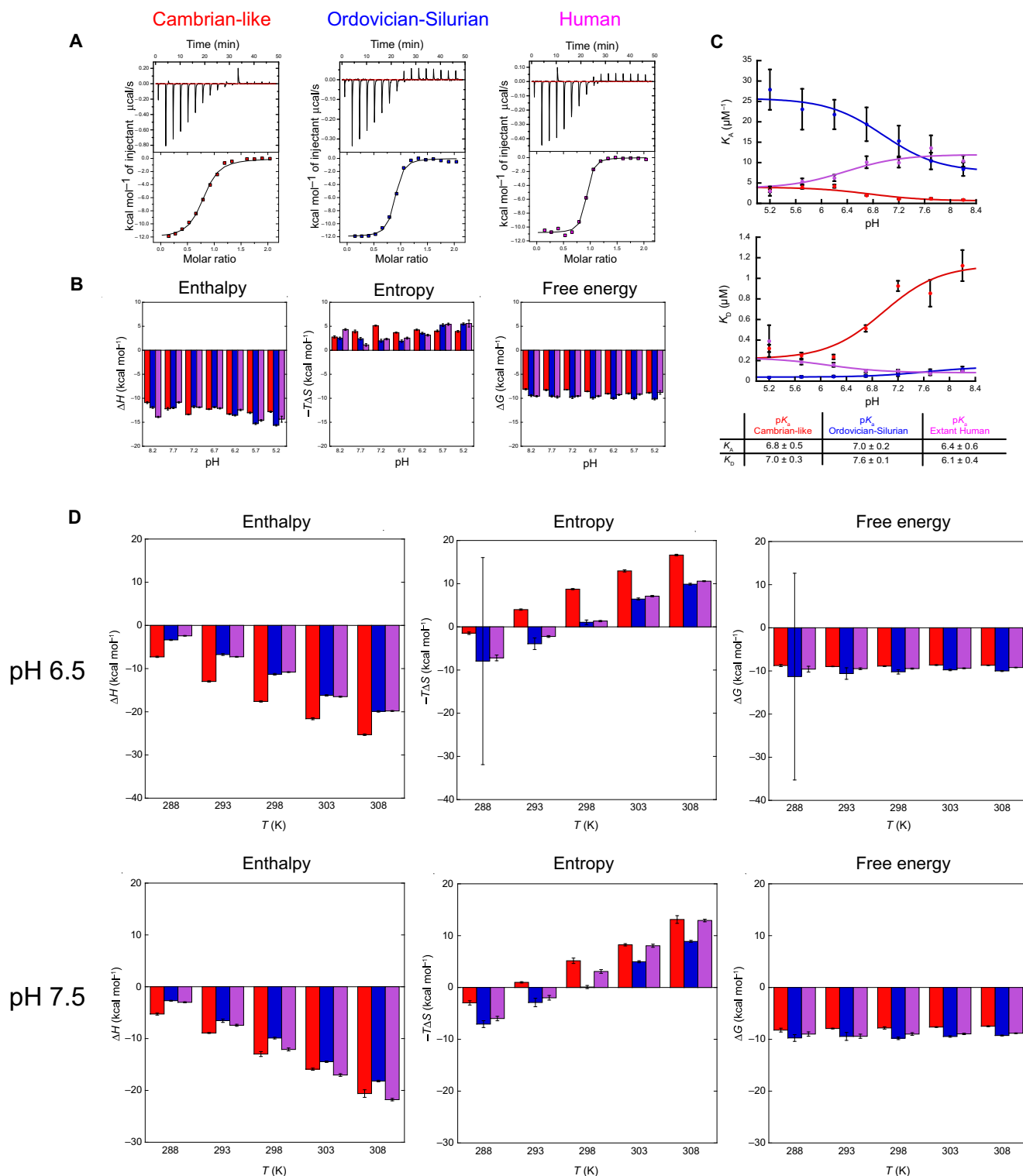
Fig. 5. The pH and temperature dependence of interactions for ancestral and extant NCBD/CID. (**A**) Examples of ITC experiments conducted at pH 7.7 and 25°C for the three different complexes. (**B**) Contribution from enthalpy (ΔH, left) and entropy ($-T\Delta S$, middle) to the total free energy (ΔG, right) at different pH values for formation of Cambrian-like (red), Ordovician-Silurian (blue), and extant human (magenta) CID/NCBD complexes, respectively. (**C**) The association (K_A, top) and dissociation (K_D, bottom) equilibrium constants plotted against pH and fitted to a two-state equation (Eq. 4 in Methods), which yielded the apparent pK_a values listed in the table. Because of the large noise in the original data for the extant human complex at pH 5.2, these points were not included in the curve fitting but are shown in the figure. (**D**) Bar diagrams showing the contribution of enthalpy (ΔH, left) and entropy ($-T\Delta S$, middle) to the free energy (ΔG, right) for formation of the respective CID/NCBD complex at different temperatures and at two different pH values (6.5 and 7.5, respectively). To determine ΔC_p, the ΔH values were plotted against temperature and analyzed according to Eq. 2 (fig. S5).

First, we investigated the pH dependence of the interaction at a constant ionic strength and temperature (25°C). Thus, ITC experiments were conducted at pH values ranging from 5.2 to 8.2 (Fig. 5 and table S1). The results show that there is a favorable enthalpic and unfavorable entropic contribution to the affinity at all pH values and for all variants (Fig. 5). Furthermore, the affinities of all complexes have a transition around pH 6.5 and reach a plateau at higher pH values (Fig. 5). The Cambrian-like complex and the Ordovician-Silurian complex displayed similar pH dependences, where the affinities of both protein complexes increased toward lower pH values, maintaining a 10-fold difference in K_A. In contrast, the extant human complex displayed an opposite pH dependence in which the affinity decreased toward lower pH values to equal that of the Cambrian-like complex around pH 5.2. K_D and K_A values were fitted to a two-state equation to estimate pK_a values for the respective parameters in the different complexes. The pH dependence of K_D reflects titrations of ionizable groups in the CID/NCBD complex, whereas the pH dependence of K_A reflects titrations in free NCBD and CID (17). The data fitted reasonably well to the model, and the fitted pK_a values for K_A and K_D were between 6.1 and 7.6 (Fig. 5). Except for one His in D/P NCBD, there are no obvious side chains with a pK_a value in this range, and the most likely basis for the pH dependences of all three complexes is a general charge effect, corroborated by the relatively broad noncooperative transitions. These effects could lead to small structural perturbations of the molten globule-like NCBD domain, which, in turn, could modulate the affinity.

Next, we investigated the thermodynamic basis for the CID/NCBD affinity by performing ITC at different temperatures and at two different pH values, 7.5 (close to the affinity plateau) and 6.5 (in the pH transition region). We found that ΔH was negative (i.e., favorable) at temperatures above 285 K, whereas the entropy contribution $-T\Delta S$ switched from being favorable (at lower temperatures) to nonfavorable between 289 and 297 K for the three complexes, both at pH 6.5 and 7.5 (Fig. 5). The calorimetric experiments show that the observed change in heat capacity upon binding ΔC_p, (i.e., the temperature dependence of ΔH) for all three complexes was identical within error at pH 6.5 (-850 to -900 cal K^{-1} mol^{-1}) and in the same range at pH 7.5 (-720 to -950 cal K^{-1} mol^{-1}; fig. S5). These numbers suggest that the contribution from the change in solvent entropy to the binding energetics (i.e., the hydrophobic effect) is substantial and similar for the three complexes (18). However, there is an offset in the values of ΔH for the Ordovician-Silurian and human complexes as compared to the Cambrian-like complex at pH 6.5 such that the enthalpic contribution, which mirrors bond strength, is weakened in the higher-affinity complexes. Instead, the thermodynamic basis for the higher affinity is a less unfavorable or even favorable entropic contribution to the overall affinity, a result that is qualitatively very similar to another IDP interaction—that between FCP1 and Rap74 (19). At pH 7.5, this offset is smaller, and all three complexes display a similar contribution from ΔH. In addition, the enthalpic contribution in the human complex is accentuated at the higher temperatures at pH 7.5, which might reflect the more well-formed Cα3. From the calorimetric experiments, we conclude that (i) the net effect of bond breaking and formation (ΔH) is favorable at supposedly physiological temperatures (>12°C); (ii) the hydrophobic effect makes a major and favorable contribution to the free energy of binding; (iii) the higher affinity observed for the Ordovician-Silurian and extant human complexes as compared to the Cambrian-like complex at pH 6.5 is due to entropic effects, which are likely not related to the hydrophobic effect since ΔC_p values are similar; and (iv) there is a delicate balance between factors contributing to affinity

as shown by, for example, the large change in ΔH for the Cambrian-like complex at pH 6.5 compared to pH 7.5 and from the pH dependences of all three complexes. Although we cannot further dissect the entropic contribution in point (iii) into changes in rotational, translational, or conformational entropy (as explained above), the thermodynamic data underscore the impact of dynamics rather than noncovalent bonds for increasing the affinity of this interaction. In addition, we confirmed by urea and thermal denaturation experiments that the ancestral Cambrian-like complex was less stable as compared to the modern Ordovician-Silurian and extant human complexes, corroborating the affinity measurements by ITC (fig. S9).

Change in frustration parallels the experimental data

The concept of minimal frustration has been successfully applied to explain protein folding, and we were therefore interested in analyzing the evolution of frustration in the coupled binding-and-folding reaction of CID/NCBD. Frustration occurs when a physical system is not able to simultaneously achieve minimum energy for every part of it (20). It has been shown that, in protein-protein interactions, frustration of regions close to the binding site often changes upon association such that it is less frustrated in the complex, thus guiding specific association (20). It has been suggested that promiscuous IDP interactions display more frustration than those of well-folded proteins (21), but little is known, especially regarding evolution of frustration in IDPs. For our three CID/NCBD complexes, we therefore calculated and analyzed the mutational frustration index, for which the Frustratometer algorithm evaluates the energy of each contact of the native structure with respect to the distribution of the energies of decoys made by substituting residues at each position (Fig. 6) (22). It is apparent that there is a significant reduction and redistribution of highly frustrated interactions in the CID domain when comparing the most ancient Cambrian-like complex with the higher-affinity younger complexes (Fig. 6). In the most ancient Cambrian-like complex, the regions around residues 1056 to 1062 of the CID domain show a distinct frustrated patch, consistent with the location of a high-mobility region. It has been previously observed that regions of high local frustration correlate with extensive subnanosecond (23) and also slower time scale dynamics (24). In the extant human complex, the amount of frustration in this region is reduced and redistributed with respect to the ancient ones. Consistent with the pH dependence of the affinity, most of the frustration changes are related to electrostatic interactions, as the changes of frustration patterns are less evident when electrostatics is not considered in the calculation (figs. S6 and S9).

DISCUSSION

Vertebrates originated during Cambrian (485 to 540 Ma), and one lineage gave rise to fishes during the subsequent Ordovician period (440 to 485 Ma). Sometime around the end of Ordovician or beginning of Silurian, a dramatic event occurred that shaped much of the following vertebrate evolution, namely, the two whole-genome duplications 1R and 2R (7, 25). This resulted in four copies of each gene and possibilities of evolutionary innovation of new functions for the encoded proteins. As the ancestral vertebrate population diversified into cartilaginous fish, bony fish, and later tetrapods including present-day amphibians, reptiles, birds, and mammals, many of the duplicated genes were subject to neo- or subfunctionalization. The genome duplications paved the way for extensive evolution of protein-protein interactions, which play a major role in cell signaling and regulation and often involve intrinsically
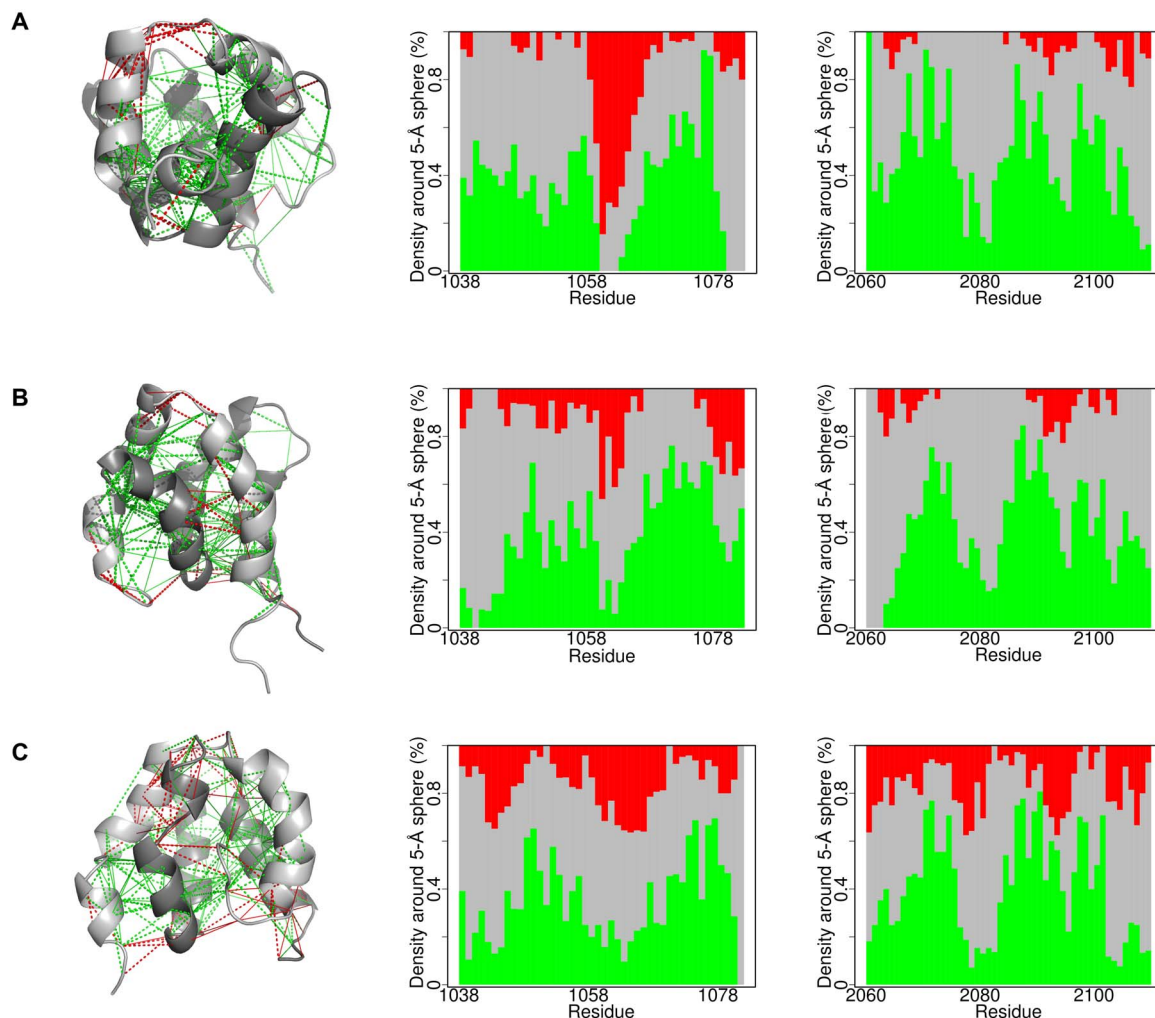
Fig. 6. Evolution of frustration in the CID/NCBD complex. The local frustration patterns were calculated for each complex using the electrostatics mode of the Frustratometer (22). Left: CID/NCBD backbones are displayed as gray ribbons (darkest gray is NCBD), direct contacts with solid lines, and water-mediated interactions with dashed lines. Minimally frustrated interactions (green) and highly frustrated contacts (red) are shown. Middle and right: The proportion of contacts within 5 Å of the Cα atom of each residue is plotted and classified according to their frustration index. (**A**) The Cambrian-like complex contains 1R CID (middle) and D/P NCBD (right). (**B**) The Ordovician-Silurian complex contains 1R CID (middle) and 1R/2R NCBD (right). (**C**) The extant human complex contains NCOA3 CID (middle) and CREBBP NCBD (right).

disordered regions. Two examples are the protein domains involved in transcriptional regulation examined here—NCBD, which is present in extant CREBBP/p300 (two sister genes were lost during evolution) and CID, which is present in NCOA1 (Src1), NCOA2 (Tif2) and NCOA3 (ACTR) (one gene was lost). The analysis of this protein-protein interaction shed light on how a functional interaction can emerge from initially weak binding between IDPs (6) and the molecular complexity of increasing affinity.

One of the fundamental differences of IDPs as compared to ordered proteins is their ability to adopt alternative bound conformations, which is advantageous when recognizing different protein ligands and which would facilitate neofunctionalization as for CID recruiting NCBD as binding partner (6). Not only the disordered C-terminal domain of p53 is a prime example of such conformational plasticity (26) but also the NCBD domain, which adopts a distinct structure upon binding to a protein domain from the interferon regulatory factor (27), as compared to binding the CID domain. However, little is known regarding the evolution of conformational plasticity and how it modulates affinity

in a protein-protein interaction. Our three historical structural and dynamic snapshots of the CID/NCBD complex provide a unique example of how increased affinity can be attained.

Overall, our findings support the notion that the high conformational plasticity of IDPs contributes to their protein-protein interactions (28). In the present case, the Cα3 of 1R CID adopts a helical conformation in the high-affinity complex with 1R/2R NCBD but not in the low-affinity complex with D/P NCBD. We speculate that the initial ancestral interaction between CID and NCBD was of lower affinity than our most ancient Cambrian-like complex yet achievable through the N-terminal leucine-rich binding motifs. Our NMR structures, NMR dynamics and thermodynamic data, and previous molecular dynamics simulations (6) illustrate how higher affinity in IDP interactions can be achieved by lowering the entropic penalty and increasing the enthalpic contribution of complex formation through a combination of specific interactions, conformational rearrangements, and redistribution of dynamics and frustration patterns. The most ancient Cambrian-like complex is more dynamic, which is probably a reflection of the fewer

specific interactions and a very high local frustration not present in the higher-affinity variants. The high-affinity Ordovician-Silurian 1R/2R complex displays rearrangements of helices along with an increased hydrophobic interface. Thus, the conformational plasticity of the interacting IDPs permits formation of new favorable interactions between Cα3 and Nα2, leading to higher affinity in the Ordovician-Silurian and the extant human complex. Why are these features not clearly reflected in a larger enthalpic contribution to binding for the high-affinity complexes? For example, the affinity of the Ordovician-Silurian complex is always 10-fold higher than the Cambrian-like complex, but the enthalpy (supposedly reflecting structure) is less favorable and the entropy (supposedly reflecting dynamics) is less unfavorable except at low pH. The answer must lie in the large and opposing overall

contributions from enthalpy and entropy and the hard-to-predict effect of water. Thus, while the overall thermodynamics of this coupled binding-and-folding reaction, which includes formation of a well-defined hydrophobic core, follows the same principles as observed in protein folding studies (29), they cannot easily be correlated to structural details. This observation is likely a reflection of a general property of IDP interactions, namely, they are plastic and thereby able to adapt to its binding partner.

The evolution of new protein-protein interactions involving intrinsically disordered regions of proteins is very common in nature as evident from virus biology. Viruses often hijack host protein-protein interactions by simply mimicking the host's disordered interaction motifs (30). Because of the high mutation rate and short generation time of viral genomes, such evolution happens on a time scale much

Table 1. Structural determination statistics.

	Most ancient	1R/2R	Extant human
NMR distance and dihedral restraints			
Distance restraints			
Total NOEs	620	700	550
Intraresidue	222	265	213
Sequential ($\lvert i - j \rvert = 1$)	197	206	168
Medium-range ($1 < \lvert i - j \rvert < 5$)	140	164	97
Long-range ($\lvert i - j \rvert > 5$)	45	55	38
Total RDC restraints ($^{15}N\text{-}^{1}H^{N}$)	91	85	79
Total dihedral angle restraints			
$^{3}J_{HNHA}$ scalar couplings	70	63	67
$^{13}C\alpha$ chemical shifts	138	142	140
Structure statistics			
Average CYANA target function value (Å^2)	7.9 ± 0.2	4.2 ± 0.30	2.2 ± 0.2
Violations			
Distance constraints (>0.5 Å)	0	3	0
Dihedral angle constraints (>5°)	0	0	0
Deviations from idealized geometry			
Bond lengths (Å)	0	0	0
Bond angles (°)	0	0	0
Impropers (°)	0	0	0
Average pairwise RMSD (Å)*			
Backbone	0.8 ± 0.2	0.6 ± 0.1	1.7 ± 0.6
Heavy atoms	1.3 ± 0.2	1.1 ± 0.1	2.2 ± 0.5
RMSD of RDC-refined structure (Å)*			
Backbone	1.3 ± 0.3	0.8 ± 0.2	1.4 ± 0.7
Heavy atoms	1.8 ± 0.2	1.3 ± 0.2	1.9 ± 0.6

*RMSD for residues 1042 to 1079 and 2065 to 2106.

shorter than that for animals. However, the general molecular principles governing emergence of new protein-protein interactions are expected to be general. Our results demonstrate how protein-protein complexes involving intrinsically disordered domains can achieve stronger binding via an intricate interplay between structural and dynamical rearrangements that cannot easily be predicted.

METHODS

Protein expression
The constructs for the CID domains contained the complementary DNA corresponding to residues 1040 to 1080 of human NCOA3, and the constructs for the NCBD domains contained residues 2062 to 2109 of human CREBBP NCBD, as described (6). All constructs contained an N-terminal expression tag (6× His and lipoyl domain), followed by a thrombin protease site and were expressed from a pRSET vector. The CID and the NCBD domains were expressed in *Escherichia coli* BL21 (DE3) pLysS cells either unlabeled (in terrific broth medium) or labeled for NMR experiments (in M9 minimal medium enriched with ^{15}N ammonium chloride and ^{13}C D-glucose) and purified, as previously described (6). Concisely, on day 1, BL21 plyS cells containing transformed plasmid were grown overnight at 37°C on a plate containing ampicillin (100 μg/ml) and chloramphenicol (35 μg/ml). On day 2, transformed cells were grown in growth media containing ampicillin (50 μg/ml) at 37°C to an optical density of 0.6. Protein expression was induced by addition of 1 mM isopropyl-β-D-thiogalactopyranoside, and cells were allowed to grow overnight at 18°C. On day 3, these cells were harvested by centrifugation and resuspended in purification buffer [20 mM tris/Cl (pH 8.0) and 200 mM NaCl]. The resuspended cells were then lysed by sonication and purified on a Ni Sepharose column (GE Healthcare). After desalting the partially pure fusion protein, ca. 0.5 units of thrombin protease per milligram of fusion protein was added to the sample and incubated at 37°C for 16 to 18 hours. The cleaved proteins were then passed through a second Ni Sepharose column, and unbound cleaved CID or NCBD was loaded on a reversed-phase C18 column equilibrated with 0.1 % trifluoroacetic acid. Bound proteins were eluted with a gradient of 0 to 70% acetonitrile in 0.1% trifluoroacetic acid. The purity and identity of the proteins were confirmed by SDS–polyacrylamide gel electrophoresis and mass spectrometry. Pure proteins were lyophilized and stored at −20°C until use. Before NMR experiments, the lyophilized samples were dissolved in a buffer containing 10 mM NaPO$_4$ (pH 6.8) and 150 mM NaCl and were dialyzed against the same buffer overnight at 25°C.

NMR spectroscopy
Protein concentrations were estimated by absorbance measurement at 205 nm for all domains and double checked at 280 nm for 1R/2R NCBD and human CREBBP NCBD, which contain one tyrosine. The protein complexes were formed by adding saturating amounts of either unlabeled CID domain (^{14}N/^{12}C) with labeled NCBD domain (^{15}N/^{13}C) or vice versa. The final concentrations of the CID/NCBD complexes in the NMR tube ranged from 300 to 500 μM of the labeled protein and approximately a twofold excess of unlabeled protein. To these samples, 0.01% NaN$_3$ and 5% D$_2$O were added. All NMR experiments were recorded on Bruker spectrometers (600, 700, and 900 MHz, ^{1}H frequencies), equipped with triple-resonance cryogenic temperature probes at 298 K. Assignment of the protein backbone was achieved by measuring two-dimensional (2D) [^{1}H,^{15}N]-HSQC (heteronuclear single-quantum coherence), 3D HNCACB, and [^{1}H,^{1}H]-NOESY-

[^{1}H,^{15}N]-HSQC experiments (31). Side chains were assigned from the following 3D experiments: [^{1}H,^{15}N]-HSQC-[^{1}H^{13}C-^{13}C^{1}H]-TOCSY (total correlation spectroscopy; 60-ms mixing time), [^{1}H,^{1}H]-NOESY-[^{1}H,^{13}C]-HSQC, [^{1}H,^{13}C]-HSQC-[^{1}H^{13}C-^{13}C^{1}H]-TOCSY (60-ms mixing time) in combination with 2D [^{1}H,^{15}N]-HSQC, and 2D [^{1}H,^{13}C]-HSQC-CT (constant time; fig. S7) (31). The φ angle–restraining $^3J_{HNHA}$ couplings were determined from a 3D HNHA-type experiment using quantitative *J* coupling intensity evolution (32). 2D and 3D in-phase and anti-phase [^{1}H-^{15}N]-HSQC–based and [^{1}H-^{15}N/^{13}C] HNCO–based experiments were used for the determination of $^1D_{HN}$ RDCs in a sample, with proteins aligned by pf1 phage. The differences between the apparent scalar couplings in the absence and presence of alignment media yielded the RDCs (fig. S8). For the collection of distance restraints, the following NOESY (NOE spectroscopy) experiments and parameters were used: 3D NOESY-[^{1}H,^{15}N]-HMQC (heteronuclear multiple-quantum coherence) and NOESY-[^{1}H,^{13}C]-HMQC with 60-ms mixing time and 128 (^{15}N or ^{13}C) × 200 (^{1}H) × 2048 (^{1}H, direct) number of points for intramolecular NOEs and a [^{14}N-^{1}H]-[^{12}C-^{1}H] double-filtered NOESY-[^{1}H,^{15}N/^{13}C]-HSQC for intermolecular NOEs (33). R_1, and $R_{1\rho}$ relaxation rates were measured with different randomized relaxation delays in an interleaved manner, with a recovery delay of 3 s using TROSY (transverse relaxation optimized spectroscopy)–based pulse schemes (34). For $R_{1\rho}$, a spin lock of 2000 Hz was used. R_2 values were then determined from $R_{1\rho}$ by correcting for the contribution of R_1. The overall tumbling time τ$_c$ was then estimated from the ratio of R_2 ($R_{1\rho}$)/R_1. [^{1}H,^{15}N]-heteronuclear NOE experiments were recorded, with a ^{1}H saturation time of 8 s with an equal recovery time for the reference experiment (34). CPMG dispersion profiles were determined from relaxation rates measured with a fixed CPMG transverse rotating-frame relaxation delay t_{CPMG} (48 ms) with ν$_{CPMG}$ frequencies of 83, 167, 250, 333, 417, 500, 583, 667, 750, 883, 917, and 1000 Hz applied in a randomized manner (35). The relaxation delay was set to 2 s. A reference experiment omitting the CPMG element was used to obtain the effective relaxation rate R_{eff} from the expression −ln (I/I_0)/t_{CPMG}, where I and I_0 are the intensities at different CPMG times and reference intensities, respectively. All experiments were processed with NMRPipe (36) and analyzed with CCPNmr (37). Curve fitting was done in MATLAB, NMRPipe, and KaleidaGraph. Errors were determined from Monte Carlo simulations. In the fast exchange limit, the effective relaxation rate as a function of ν$_{CPMG}$ can be described by the following equation (13)

$$R_{2eff} =$$

$$R_{20} + \left((p_A p_B \Delta\omega^2)/k_{ex}(1 - 4t_{CPMG}/k_{ex}) \tanh\left(k_{ex}/(4t_{CPMG})\right)\right) \quad (1)$$

where, R_{20} is the transverse relaxation rate in the absence of exchange, p$_A$ and p$_B$ are the populations of the exchanging states, $\Delta\omega$ is the chemical shift difference between the exchanging states, k_{ex} is the sum of the forward and reverse exchange rate constants, and τ$_{cp}$ are the variable CPMG times. Chemical shift order parameters (S^2) were predicted by submitting experimentally derived shifts to a web-based program from the Wishart laboratory (38). R_{ex} was calculated by taking the difference between R_2 (R_{2eff}) without and with a CPMG radio pulse applied at a frequency of 1000 Hz (fig. S9).

Structure calculations
Structure calculations were done using the CYANA 3.97 package in two steps. First, the NOESY cross peaks were converted into upper distance restraints (fig. S7) following an automated process in CYANA. These

distance restraints, together with ϕ/ψ dihedral angles determined from Cα chemical shifts and $^3J_{HNHA}$ (fig. S8), were used as input for the initial structure calculations. The structures were calculated with 200,000 torsion angle dynamics steps for 100 conformers, starting from random torsion angles by simulated annealing. Initially, the structures of the CID and NCBD domains were determined separately without the intermolecular NOEs. In a second step, the calculations were repeated using the intermolecular NOE-derived distance restraints (table S1) and a linker with a weak harmonic well potential with a bottom width of 1.2 Å that kept the two domains together (39). The separate structures calculated in step 1 were used as the starting structures. For refinement and cross-validation, calculations were repeated with $^1D_{HN}$ RDCs (fig. S8, F to K). The axially symmetric and rhombic components of the alignment tensor were optimized by an iterative process in CYANA in three steps. For representation and analysis, the 20 conformers with the lowest target function values were selected. The structural statistics, together with all input data for the structure calculations, are presented in Table 1 and figs. S8 and S9. The total SASA was determined using the program PyMol from the domains in the complex.

Isothermal titration calorimetry

ITC binding experiments were performed on an iTC200 (Malvern Instruments) and fitted using in-built software. Before performing the ITC experiments, all protein samples were dialyzed overnight in a buffer containing 20 mM sodium phosphate at pH 6.5 or 7.5. The protein samples were then filtered, and their concentrations were determined using absorbance at 205 nm. The ITC experiments were done at different temperatures (288, 293, 298, 303, and 308 K). Typically, 20 injections of 2 μl each of 300 μM CID domain solution were titrated into 280 μl of 30 μM NCBD domains. All samples were allowed to equilibrate for about 10 min at the experimental temperature before the actual measurements were initiated. For the determination of ΔC_p, both ΔH and ΔS were plotted as a function of temperature following Eqs. 2 and 3.

$$\Delta H(T) = \Delta C_p(T) + \Delta_x H \qquad (2)$$

$$\Delta S(T) = \Delta C_p \ln(T) + \Delta_x S \qquad (3)$$

ΔC_p is the specific heat capacity at constant pressure. $\Delta_x H$ and $\Delta_x S$ are the temperature-independent contributions of enthalpy and entropy, respectively. The ITC measurements for pH dependence of affinity were carried out essentially as above, with the following differences: The protein samples were dialyzed against 20 mM sodium phosphate buffer (pH 8.2, 7.7, 7.2, 6.7, and 6.2) or, alternatively, 20 mM sodium acetate buffer (pH 5.7 and 5.2), and the ionic strength was adjusted with NaCl to 58 mM for all buffers. All measurements were conducted at 298 K. The measurement for the extant human complex at pH 5.2 was repeated four times because of problems with noise in the isotherms, and these data points were excluded from the fitting of pK_a values. The data were fitted to a two-state equation (Eq. 4) to extract apparent pK_a values for the association (K_A) and dissociation (K_D) constants for the respective protein complex.

$$K = (K_{protonated}10^{-pH} + K_{deprotonated}10^{-pKa})/(10^{-pKa} + 10^{-pH}) \qquad (4)$$

Stability measurements

Urea and thermal denaturation experiments were performed on a JASCO J-1500 circular dichroism (CD) spectrophotometer, and the data were fitted using the program KaleidaGraph. All lyophilized protein samples were dissolved in a buffer containing 10 mM sodium phosphate (pH 6.8). For each complex, 40 μM of the CID domain was mixed with 15 μM NCBD domain. The temperature was set to 298 K for the urea denaturation experiments. The temperature was varied between 278 and 363 K for the thermal denaturation. The ellipticity at 222 nm was monitored for both measurements. The data were then fitted to the standard equations describing two-state unfolding by chemical denaturant and temperature, respectively (17).

Frustration index calculations

The frustration index was calculated for each contact in the complex using the Frustratometer, as described (22). A contact with an index of ≥0.78 is classified as minimally frustrated, contributing favorably to protein stability and affinity. On the other hand, interactions with a frustration index of ≤−1 are labeled as highly frustrated, and most substitutions at that location would be more favorable. In between these extremes, the interactions are labeled as neutral. Interactions are divided into classes according to this index and color coded in the figures. Considering that the CID/NCBD complexes are charged under the pH conditions used for the experiments, we computed the local frustration patterns using the electrostatics module of the Frustratometers, with an electrostatics k = 4.15 (22), and compared the results with calculations not considering the electrostatics.

SUPPLEMENTARY MATERIALS

Fig. S1. Structural changes taking place and chemical shift restraints.
Fig. S2. Strips from [^{1}H-^{1}H]-NOESY-[^{1}H-^{15}N]-HSQC spectra showing resonances emanating from residues 1074 and 1075 of the respective CID domains.
Fig. S3. Strips from [^{1}H-^{1}H]-NOESY [^{1}H-^{15}N]-HSQC spectra showing resonances emanating from residues 1076 and 1077 of the CID domains.
Fig. S4. ^{15}N relaxation data for the bound CID and NCBD domains and amide secondary chemical shifts of the complexes.
Fig. S5. Backbone subnanosecond motions and ΔC_p for the three CID/NCBD complexes.
Fig. S6. Frustration in the CID/NCBD complexes without electrostatics.
Fig. S7. [^{1}H-^{15}N]-HSQC correlation spectra for bound and free CID and NCBD domains and the number of restraints in the structure determination as a function of amino acid residue.
Fig. S8. Restraints used for structure calculations and refinement.
Fig. S9. Urea and temperature denaturation experiments and Coulomb surfaces of the ancestral and extant CID/NCBD complexes.
Table S1. ITC parameters and intermolecular NOEs.

REFERENCES AND NOTES

1. C. Leo, J. D. Chen, The SRC family of nuclear receptor coactivators. *Gene* **245**, 1–11 (2000).
2. M. J. Harms, J. W. Thornton, Evolutionary biochemistry: Revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* **14**, 559–571 (2013).
3. A. N. McKeown, J. T. Bridgham, D. W. Anderson, M. N. Murphy, E. A. Ortlund, J. W. Thornton, Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell* **159**, 58–68 (2014).
4. S. A. Lim, K. M. Hart, M. J. Harms, S. Marqusee, Evolutionary trend toward kinetic stability in the folding trajectory of RNases H. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13045–13050 (2016).
5. V. Nguyen, C. Wilson, M. Hoemberger, J. B. Stiller, R. V. Agafonov, S. Kutter, J. English, D. L. Theobald, D. Kern, Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science* **355**, 289–294 (2017).
6. G. Hultqvist, E. Åberg, C. Camilloni, G. N. Sundell, E. Andersson, J. Dogan, C. N. Chi, M. Vendruscolo, P. Jemth, Emergence and evolution of an interaction between intrinsically disordered proteins. *eLife* **6**, e16059 (2017).
7. T. Hughes, D. A. Liberles, Whole-genome duplications in the ancestral vertebrate are detectable in the distribution of gene family sizes of tetrapod species. *J. Mol. Evol.* **67**, 343–357 (2008).

8. N. Tjandra, A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–1114 (1997).

9. A. Mascioni, G. Veglia, Theoretical analysis of residual dipolar coupling patterns in regular secondary structures of proteins. *J. Am. Chem. Soc.* **125**, 12520–12526 (2003).

10. S. J. Demarest, M. Martinez-Yamout, J. Chung, H. Chen, W. Xu, H. J. Dyson, R. M. Evans, P. E. Wright, Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **415**, 549–553 (2002).

11. N. Salvi, A. Abyzov, M. Blackledge, Multi-timescale dynamics in intrinsically disordered proteins from NMR relaxation and molecular simulation. *J. Phys. Chem. Lett.* **7**, 2483–2489 (2016).

12. N. E. Zhou, B. Y. Zhu, B. D. Sykes, R. S. Hodges, Relationship between amide proton chemical shifts and hydrogen bonding in amphipathic .alpha.-helical peptides. *J. Am. Chem. Soc.* **114**, 4320–4326 (1992).

13. A. G. Palmer III, NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* **104**, 3623–3640 (2004).

14. A. Abyzov, N. Salvi, R. Schneider, D. Maurin, R. W. H. Ruigrok, M. R. Jensen, M. Blackledge, Identification of dynamic modes in an intrinsically disordered protein using temperature-dependent NMR relaxation. *J. Am. Chem. Soc.* **138**, 6240–6251 (2016).

15. M. S. Marlow, J. Dogan, K. K. Frederick, K. G. Valentine, A. J. Wand, The role of conformational entropy in molecular recognition by calmodulin. *Nat. Chem. Biol.* **6**, 352–358 (2010).

16. S.-R. Tzeng, C. G. Kalodimos, Protein activity regulation by conformational entropy. *Nature* **488**, 236–240 (2012).

17. A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Macmillan, 1999).

18. R. L. Baldwin, Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8069–8072 (1986).

19. C. W. Lawrence, S. Kumar, W. G. Noid, S. A. Showalter, Role of ordered proteins in the folding-upon-binding of intrinsically disordered proteins. *J. Phys. Chem. Lett.* **5**, 833–838 (2014).

20. D. U. Ferreiro, E. A. Komives, P. G. Wolynes, Frustration, function and folding. *Curr. Opin. Struct. Biol.* **48**, 68–73 (2018).

21. P. Jemth, X. Mu, Å. Engström, J. A. Dogan, A frustrated binding interface for intrinsically disordered proteins. *J. Biol. Chem.* **289**, 5528–5533 (2014).

22. R. G. Parra, N. P. Schafer, L. G. Radusky, M.-Y. Tsai, A. B. Guzovsky, P. G. Wolynes, D. U. Ferreiro, Protein Frustratometer 2: A tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res.* **44**, W356–W360 (2016).

23. I. Lindström, J. Dogan, Dynamics, conformational entropy, and frustration in protein-protein interactions involving an intrinsically disordered protein domain. *ACS Chem. Biol.* **13**, 1218–1227 (2018).

24. B. Fuglestad, P. M. Gasper, J. A. McCammon, P. R. L. Markwick, E. A. Komives, Correlated motions and residual frustration in thrombin. *J. Phys. Chem. B* **117**, 12857–12863 (2013).

25. S. Kuraku, A. Meyer, S. Kuratani, Timing of genome duplications relative to the origin of the vertebrates: Did cyclostomes diverge before or after? *Mol. Biol. Evol.* **26**, 47–59 (2008).

26. C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Yang, V. N. Uversky, A. K. Dunker, Flexible nets: Disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* **9** (suppl. 1), S1 (2008).

27. B. Y. Qin, C. Liu, H. Srinath, S. S. Lam, J. J. Correia, R. Derynck, K. Lin, Crystal structure of IRF-3 in complex with CBP. *Structure* **13**, 1269–1277 (2005).

28. P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).

29. C. M. Johnson, A. R. Fersht, Protein stability as a function of denaturant concentration: The thermal stability of barnase in the presence of urea. *Biochemistry* **34**, 6795–6804 (1995).

30. N. E. Davey, G. Travé, T. J. Gibson, How viruses hijack cell regulation. *Trends Biochem. Sci.* **36**, 159–169 (2011).

31. J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, N. J. Skelton, *Protein NMR Spectroscopy: Principles and Practice* (Academic Press, ed. 2, 2007).

32. G. W. Vuister, A. Bax, Measurement of four-bond H^N-H^α J-couplings in staphylococcal nuclease. *J. Biomol. NMR* **4**, 193–200 (1994).

33. C. Eichmüller, W. Schüler, R. Konrat, B. Kräutler, Simultaneous measurement of intra- and intermolecular NOEs in differentially labeled protein-ligand complexes. *J. Biomol. NMR* **21**, 107–116 (2001).

34. N.-A. Lakomek, J. Ying, A. Bax, Measurement of ^{15}N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR* **53**, 209–221 (2012).

35. D. Ban, A. Mazur, M. G. Carneiro, T. M. Sabo, K. Giller, L. M. I. Koharudin, S. Becker, A. M. Gronenborn, C. Griesinger, D. Lee, Enhanced accuracy of kinetic information from CT-CPMG experiments by transverse rotating-frame spectroscopy. *J. Biomol. NMR* **57**, 73–82 (2013).

36. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

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

38. M. V. Berjanskii, D. S. Wishart, A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* **127**, 14970–14971 (2005).

39. C. N. Chi, B. Vögeli, S. Bibow, D. Strotz, J. Orts, P. Güntert, R. Riek, A structural ensemble for the enzyme cyclophilin reveals an orchestrated mode of action at atomic resolution. *Angew. Chem. Int. Ed.* **54**, 11657–11661 (2015).

Citation: P. Jemth, E. Karlsson, B. Vögeli, B. Guzovsky, E. Andersson, G. Hultqvist, J. Dogan, P. Güntert, R. Riek, C. N. Chi, Structure and dynamics conspire in the evolution of affinity between intrinsically disordered proteins. *Sci. Adv.* **4**, eaau4130 (2018).

Science Advances

Structure and dynamics conspire in the evolution of affinity between intrinsically disordered proteins

Per Jemth, Elin Karlsson, Beat Vögeli, Brenda Guzovsky, Eva Andersson, Greta Hultqvist, Jakob Dogan, Peter Güntert, Roland Riek and Celestine N. Chi

ARTICLE TOOLS	http://advances.sciencemag.org/content/4/10/eaau4130
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2018/10/22/4.10.eaau4130.DC1
REFERENCES	This article cites 37 articles, 5 of which you can access for free http://advances.sciencemag.org/content/4/10/eaau4130#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service