UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1870

Evolution and Binding Mechanisms of Intrinsically Disordered Proteins

ELIN KARLSSON

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022

ISSN 1651-6206
ISBN 978-91-513-1604-8
URN urn:nbn:se:uu:diva-485228

Dissertation presented at Uppsala University to be publicly examined in A1:107a, Biomedicinskt centrum, Husargatan 3, Uppsala, Monday, 14 November 2022 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Kaare Teilum (Department of Biology, University of Copenhagen).

Abstract
Karlsson, E. 2022. Evolution and Binding Mechanisms of Intrinsically Disordered Proteins. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1870. 53 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1604-8.

Intrinsically disordered proteins (IDPs) make up a considerable fraction of the proteome in eukaryotic organisms. These proteins often act as hubs in interaction networks, harbouring multiple interaction with other proteins, and thus evolution has to walk a tightrope to accommodate new interactions while maintaining the previously established interactions. The ability to accommodate multiple ligands with high specificity is one of the fascinating properties of IDPs, and the molecular details of how this is achieved throughout evolution are largely unknown. The nuclear co-activator binding domain (NCBD) and CREBBP-interacting domain (CID) are intrinsically disordered domains of two transcriptional co-activator proteins. This interaction constitutes one of the earliest examples of a binding reaction where two binding partners fold synergistically upon binding. Previous phylogenetic analysis showed that NCBD is evolutionarily older than CID, which likely emerged after the divergence of the deuterostome and protostome clades of the animal kingdom. When NCBD adapted to the new ligand CID, the affinity of this interaction increased 10-20-fold, while the affinity for some other NCBD ligands were maintained. My thesis work has largely focused around understanding the evolutionary adaptation of the NCBD-CID complex. I have characterised a reconstructed ancestral NCBD-CID complex with respect to structure, folding and binding mechanism and compared these properties to those of the present-day human complex. The results show that the structure and disordered properties of NCBD and CID, as well as their overall binding mechanism, have been moderately conserved throughout evolution. Small differences in the binding mechanism and compactness of the complexes were observed between the most ancestral and present-day human complexes, indicating a somewhat malleable protein-protein interaction that allows for fine-tuning of biophysical properties when new ligands are adopted. Furthermore, I have investigated the impact of disordered regions flanking the binding interface of present-day human CID, using stopped-flow fluorimetry. The disordered regions contributed to an increased affinity to NCBD, although no additional structure was formed upon binding. Ionic strength-dependence curves of the obtained kinetic parameters showed that electrostatic interactions likely do not contribute to the increase in affinity mediated by the disordered flanking regions. These results demonstrate how disordered regions flanking the binding interface regions can contribute to affinity, and highlights the importance of including larger parts of proteins when conducting studies of proteins *in vitro*.

Keywords: Intrinsically disordered proteins, Protein evolution, Binding kinetics, Binding mechanisms, Biophysical characterisation

Elin Karlsson, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.

ISSN 1651-6206
ISBN 978-91-513-1604-8
URN urn:nbn:se:uu:diva-485228 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-485228)

Till Fredrik, Ralf & Lillebror

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. Jemth, P., **Karlsson, E**., Vögeli, B., Guzovsky, B., Andersson, E., Hultqvist, G., Dogan, J., Günthert, P., Riek, R., Chi, C. (2018) Structure and dynamics conspire in the evolution of affinity between intrinsically disordered proteins. *Sci. Adv*, 4(10), eaau4130.

II. **Karlsson, E**., Paissoni, C., Erkelens, A M., Tehranizadeh, Z A., Sorgenfrei, F., 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(51), 17698–17712.

III. **Karlsson, E**., Lindberg, A., Jemth, P. (2020). High affinity between CREBBP/p300 and NCOA evolved in vertebrates. *Protein Sci*. 29(7), 1687–1691.

IV. **Karlsson, E**., Sorgenfrei, F., Andersson, A., Dogan, J., Jemth, P., Chi, C. (2022). The dynamic properties of a nuclear coactivator binding domain are evolutionarily conserved. *Commun. Biol*. 5(286).

V. **Karlsson, E**., Schnatwinkel, J., Paissoni, C., Andersson, E., Herrmann, C., Camilloni, C., Jemth, P. (2022). Disordered regions flanking the binding interface modulate affinity between CBP and NCOA. *J. Mol. Biol*. 434(13), 167643.

List of additional papers not included in the thesis

Karlsson, E., Andersson, E., Dogan, J., Gianni, S., Jemth, P., Camilloni, C. (2018). A Structurally Heterogeneous Transition State Underlies Coupled Binding and Folding of Disordered Proteins. *J. Biol. Chem*. 294(4), 1230–1239.

Karlsson, E., Andersson, E., Jones, N C., Vrønning Hoffmann, S., Jemth, P., Kjaergaard, M. (2019). Coupled Binding and Helix Formation Monitored by Synchrotron-Radiation Circular Dichroism. *Biophys. J*. 117(4), 729–742.

Laursen, L., **Karlsson**, **E**., Gianni, S., Jemth, P. (2020). Functional Interplay Between Protein Domains in a Supramodular Structure Involving the Postsynaptic Density Protein PSD-95. *J. Biol. Chem*. 295(7), 1992–2000.

Karlsson, E. & Jemth, P. (2021). Kinetic Methods of Deducing Binding Mechanisms Involving Intrinsically Disordered Proteins BT - Protein-Ligand Interactions: Methods and Applications. in (eds. Daviter, T., Johnson, C. M., McLaughlin, S. H. & Williams, M. A.) Springer US. 2263, 105-133.

Contents

Abbreviations

ACTR	Activator of Thyroid and Retinoid Receptor
CD	Circular Dichroism
CREB	Cyclic Adenosine Monophosphate Response
CBP	CREB Binding Protein
CREBBP	CREB Binding Protein
CID	CREBBP Interacting Domain
HAT	Histone Acetyl Transferase
IDP	Intrinsically Disordered Protein
IRF3	Interferon Regulatory Factor 3
ITC	Isothermal Titration Calorimetry
NCBD	Nuclear Co-Activator Binding Domain
NCOA	Nuclear Receptor Co-Activator
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
mRNA	Messenger RNA
SRC1	Steroid Receptor Coactivator 1
SLiM	Short Linear Motif
TIF2	Transcriptional Mediator/Intermediary Factor 2

Introduction

Since the emergence of modern protein biochemistry in the mid 20th century and until very recent years, the consensus belief in the field was that a well-defined protein structure was a pre-requisite for protein function[1]. This narrow view on the world of proteins, commonly referred to as the "structure-function paradigm", made sense in a time when the majority of the proteins that had been subjected to structural analysis were enzymes and abundant proteins from blood, which are typical examples of well-folded proteins that rely on their structure for function[2]. When cases of functional proteins that lacked stable structure were reported in the scientific literature, they were initially met with scepticism[3–6]. Today, the functional importance of such proteins is well-known and recent estimates suggest that nearly half of all proteins in the human proteome contain stretches of unstructured regions longer than 30 residues[7]. These proteins are collectively referred to as intrinsically disordered proteins (IDPs).

Two protein domains that belong to the IDP category are the nuclear co-activator binding domain (NCBD) from the large co-activator CREB-binding protein (CREBBP/p300), and the CREBBP-interacting domain (CID), a domain within the nuclear receptor co-activator (NCOA) paralogues. These protein domains form a high-affinity interaction in present-day human[8]. My thesis projects have revolved around these protein domains: their interaction and the impact of evolution on their biophysical properties.

I have studied ancestral as well as present-day variants of these proteins with various biophysical methods, in order to characterize their structure, affinity, binding mechanism and folding stability. Furthermore, I have investigated the impact of disordered regions that are flanking the binding interface of CID in the present-day human complex.

The most important conclusions from this thesis work are that evolution have preserved the overall disordered properties, binding mechanism and structure of NCBD and CID, but allowed for small differences that can act to fine-tune binding affinity.

Furthermore, this thesis work has shown that disordered flanking regions play a role in modulating the affinity of the human NCBD-CID interaction. This example could represent a general phenomenon among disordered flanking regions in protein-protein interactions and implies that the impact from disordered flanking regions should be considered when studies on isolated protein domains are conducted.

In the following passages, I am giving a general introduction to the field of IDPs as well as to the utilization of different biophysical methods in studies of their structure, kinetics and binding mechanisms.

Intrinsically disordered proteins and the folding energy landscape

When considering their native fold, proteins are commonly classified as either folded (i.e proteins that have a well-defined structure) or disordered (i.e proteins that lack a well-defined structure, referred to as IDPs). This categorical division of proteins as either folded or disordered is somewhat unrealistic, as all proteins undergo conformational fluctuations to some degree[9,10]. Thus, proteins are better described as being on a disorder-order continuum where highly disordered proteins on this continuum experience large-scale conformational fluctuations as opposed to more rigid proteins that only undergo minor fluctuations[11].

The IDP category can also be divided into more or less dynamic proteins, and encompasses proteins ranging from fully disordered ones to those possessing substantial secondary and tertiary structure content[12,13]. So, are there any common characteristics among IDPs that can be used to separate them from other proteins? In order to clarify this from a thermodynamic perspective, the theoretical folding energy landscape of proteins can be used as an explanatory tool. A typical folding energy landscape for a globular domain, for instance, is funnel-shaped and appears relatively smooth. Stable intermediates along the folding pathway may or may not be present and the folded state resides at the bottom of a deep well with few other conformational states energetically accessible (Figure 1a). In contrast to this, IDPs are typically characterized by large-scale conformational fluctuations and sampling of multiple conformations. The folding energy landscape of such a protein is normally described as being “rugged”, with multiple accessible states of equal or nearly equal stability (Figure 1b). The ruggedness of the folding energy landscape of IDPs appears correlated with a lower fraction of hydrophobic residues, which are known to be important drivers in folding into compact structures. Instead, the amino acid sequences of IDPs are enriched in charged and polar residues[14]. The ruggedness of the folding landscape also has important implications for the interactions that IDPs make with other biomolecules, which is discussed in the following section.

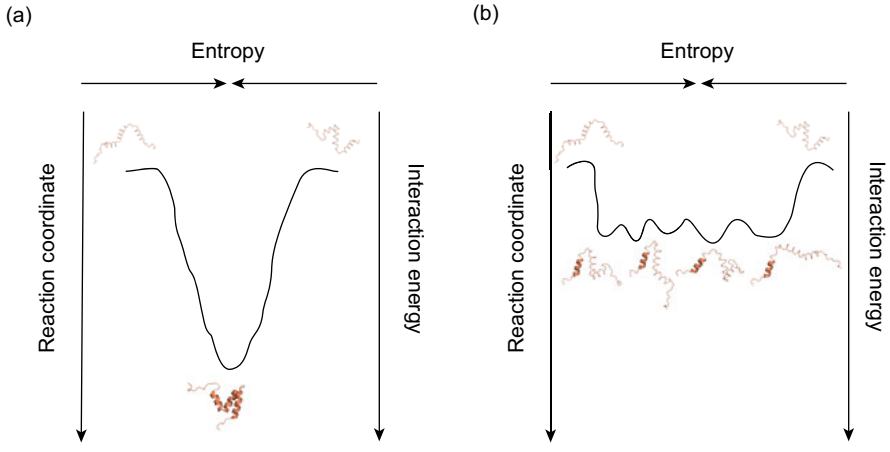

Figure 1. The folding energy landscape of different proteins on the disorder-order continuum. (a) The folding energy landscape of a structurally well-defined globular domain. The energy landscape is smooth and no populated folding intermediates are present. The deep and narrow well indicates that the folded protein structure is highly stable and that the protein structure fluctuates between a few states with very similar energy. (b) A hypothetical example of the energy landscape of an intrinsically disordered protein. The protein rapidly interconverts between different protein structures and exists as an ensemble of multiple protein states that are present at the same time.

Binding reactions involving intrinsically disordered proteins

The regulation of cellular processes takes place entirely at the level of interactions between biological molecules, which form a complex network commonly referred to as the "interactome". IDPs are frequently involved in regulatory functions in the cell and contribute to the complexity of the interactome by several mechanisms[15,16]: 1) They expose short linear motifs (SLiMs), which enable interactions with multiple targets[17]. 2) They present sites for post-translational modifications, which modulate affinities for targets and thereby regulate the flux of different signalling cascades[18,19]. 3) Genes encoding IDPs are more often subjected to splicing at the pre-mRNA level, resulting in multiple protein variants from a single gene[20]. These features provide spatial and temporal control over engagement in protein-protein interactions and has likely caused IDPs to become drivers in the development of complex multicellular organisms[21].

The link between malfunctional IDPs and severe diseases such as cancer and neurodegenerative diseases[22] has urged scientists to investigate the binding reactions in which IDPs participate. It has turned out that many IDPs fold upon binding to their partner, a phenomenon which is termed coupled binding and folding[11]. Somewhat simplified, this can occur according to either one of the

following two scenarios: 1) Folding occurs before binding (conformational selection; Figure 2a). 2) Binding occurs before folding (induced fit; Figure 2b).

In the conformational selection scenario, the IDP is interconverting between different conformations in the unbound state and folding into one of these binding-competent conformations precedes binding to the ligand. This model suggests that the IDP is only binding competent when folded in a correct conformation and thus represents an extension of the old structure-function paradigm. In the induced fit scenario, on the other hand, the IDP forms initial contacts with the partner protein in the unfolded state and this "triggers" the transition to the final (folded) bound state. There have been speculations about whether the binding of the IDP in the disordered state poses any functional advantages for the protein, such as an increased association rate constant[23]. This originates from the proposal of a "fly-casting mechanism" which is based on the notion that the disordered protein has a larger capture radius and thus a higher probability of forming early interactions with the partner protein which can guide the protein to its right conformation[24]. IDPs are likely to utilize a combination of these binding mechanisms and the flux through either one of the pathways is determined by the respective microscopic rate constants and the concentrations of protein (Figure 2c)[25].

Nevertheless, in order to distinguish between induced fit, conformational selection and other mechanisms for binding, a kinetic method is required which allows for monitoring build-up and break-down of the molecular species involved in the binding reaction. Structural information regarding the conformations sampled by the IDP as well as data from measurements performed under equilibrium can aid in the determination of a plausible model, but is alone not enough to imply a model for the binding mechanism.

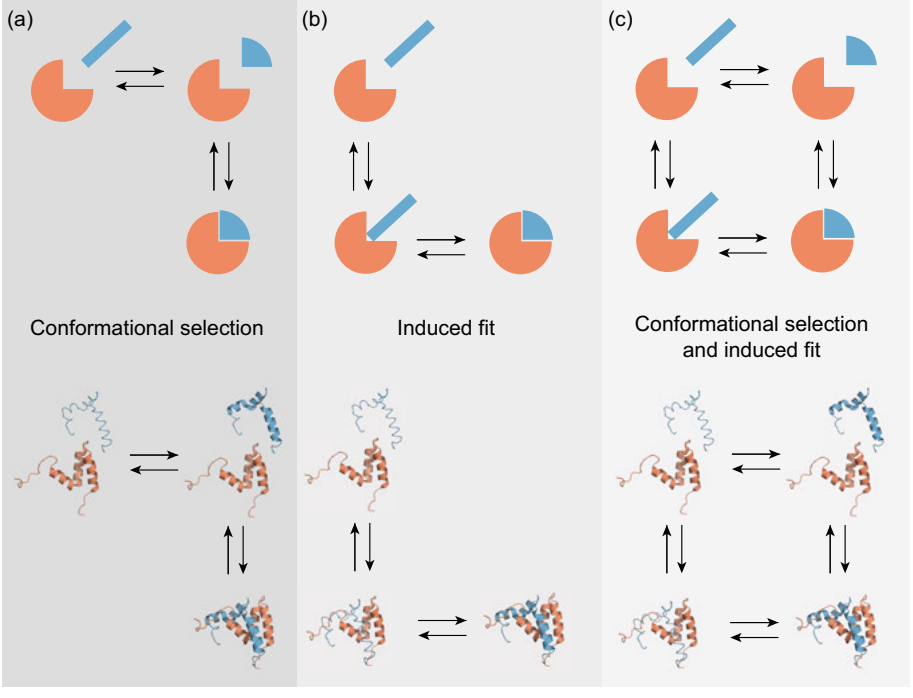

Figure 2. Different mechanisms of binding that folding-upon-binding could follow. In order to simplify the concepts, the examples in the figure are extreme and unlikely to happen in reality. (a) Conformational selection or folding before binding. The disordered binding partner (blue) first folds into a more binding competent structure and then binds to the folded target protein (orange). (b) The induced fit mechanism in which the disordered protein first forms a complex with the binding partner and then folds into the final structure. (c) A mixture of the conformational selection and induced fit models. The binding reaction can occur via both routes and the microscopic rate constants for each step will determine the flux of each pathway.

NCBD and CID – a paradigmatic protein-protein interaction

The interaction between the intrinsically disordered domains NCBD and CID constitutes an archetype example of a coupled binding and folding reaction. It was one of the earliest examples of a binding reaction where two proteins fold synergistically upon binding[8]. NCBD is a domain within CREBBP and its paralog p300. CREBBP/p300 act as a hub in transcriptional regulation and is composed of several structured domains that mediates interactions with transcription factors and transcriptional co-activators (Figure 3a). CREBBP/p300 also harbours a histone acetyl transferase (HAT) domain which modulates expression of target genes by modification of chromatin[26]. CID is a domain within proteins from the NCOA family, which comprises three paralogs in human: NCOA1/SRC1, NCOA2/TIF2 and NCOA3/ACTR (Figure 3b). The

NCOA proteins function as recruiters of transcriptional co-regulators to the transcription machinery and are important in the regulation of nuclear receptor target genes. Apart from participating in critical cellular processes, the interaction between NCBD and CID constitute an interesting model system for studies of IDP interactions.

NCBD exhibits both typical and atypical behaviour for a molten globule protein and has hence been referred to as "molten-globule-like"[8,27–30]. This domain interacts with multiple transcriptional co-regulators and transcription factors; apart from the NCOA proteins these are for instance p53 and interferon regulatory factor 3 (IRF3)[26]. NCBD also poses a target for viral hijacking during viral infections[31]. In the unbound state, NCBD mainly populates a structural state which is topologically highly similar to the structure when bound to CID and p53[8,30,32] (Figure 3c). A small fraction of the NCBD population resides in a state that resembles the IRF3-bound structure[27,33]. CID is largely disordered in the unbound state but transiently populates helical structure, in particular in the region corresponding to the N-terminal helix in the bound structure[34,35] (Figure 3c). Furthermore, this helix has attracted attention in the context of pre-formed secondary structure elements that speed up association in IDP binding reactions. Modulation of helicity of the N-terminal helix correlates strongly with affinity and affect both association and dissociation rate constants[36].

NCBD and CID interact in a coupled binding and folding reaction in which CID forms 2-3 helices with an extended interface area with NCBD[8] (Figure 3c). Kinetic studies of the NCBD-CID interaction using ensemble and single-molecule methods have revealed a quite complex reaction mechanism with numerous populated intermediates under different conditions[37–39].

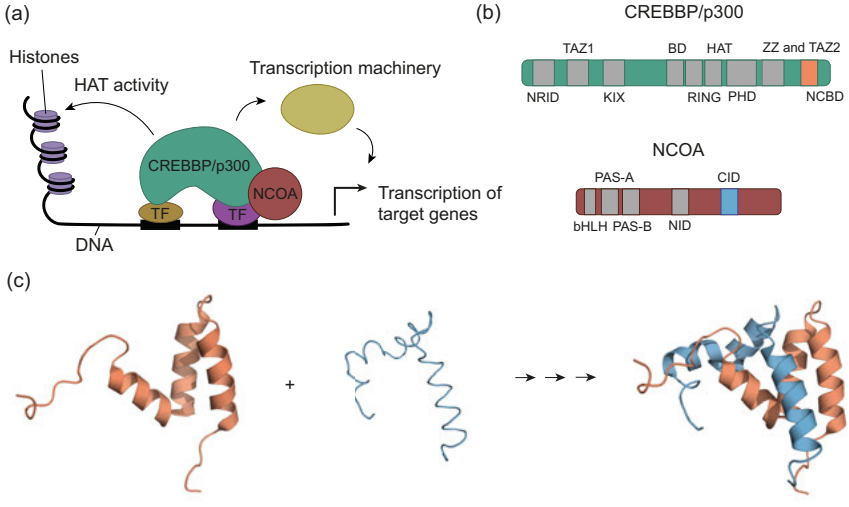

Figure 3. The transcriptional co-activating function of CREBBP/p300 and NCOA along with their domain architecture and binding reaction of the isolated NCBD and CID domains. (a) CREBBP/p300 interacts with several transcription factors (abbreviated as TF in the figure) either directly or by recruiting other transcriptional co-regulators such as NCOA, which mediates interaction with nuclear receptors. Nuclear receptors are ligand-dependent transcription factors that bind to specific sites on DNA. CREBBP/p300 harbours a histone acetyltransferase (HAT) domain which remodels chromatin and increases transcriptional activity at the DNA target site. By recruitment of the transcriptional machinery, nuclear receptor target genes are transcribed upon activation by ligand binding. (b) The domain architecture of CREBBP/p300 (top) and NCOA (bottom). The NCOA paralogues interact with nuclear receptors directly through the nuclear receptor interaction domain (NID) and interacts with CREBBP/p300 through the CREBBP-interaction domain (CID). (c) The structure of free NCBD (orange) and a model of free CID (blue). NCBD and CID engage in a synergistic folding upon binding reaction, in which CID undergoes the most dramatic increase in alpha-helical content. The structure of free NCBD was created in PyMol with the PDB entry 2KKJ[27] and the structure of the NCBD-CID complex with entry 1KBH[8]. The structure of free CID is a hypothetical model.

Evolution of the NCBD-CID interaction

NCBD has evolved to interact with multiple binding partners. Previous phylogenetic analysis suggested that NCBD is evolutionarily older than CID, which likely emerged in the proteome after the divergence of the deuterostome and protostome lineages[35]. Ancestral sequence reconstruction was used previously to reconstruct sequences of NCBD and CID from different evolutionary time points[35]. This showed that the earliest NCBD variant that could be reconstructed with decent confidence, from the common ancestor of deuterostomes and protostomes, displayed a 10-20-fold lower affinity for CID as compared to the affinity of the present-day human NCBD-CID complex[35]. An NCBD-

CID complex from the time of the whole genome duplications, which occurred in a vertebrate ancestor around 450 million years ago, displayed a similar affinity to the human complex, indicating that this interaction was well-established at this evolutionary time point (Figure 4)[35].

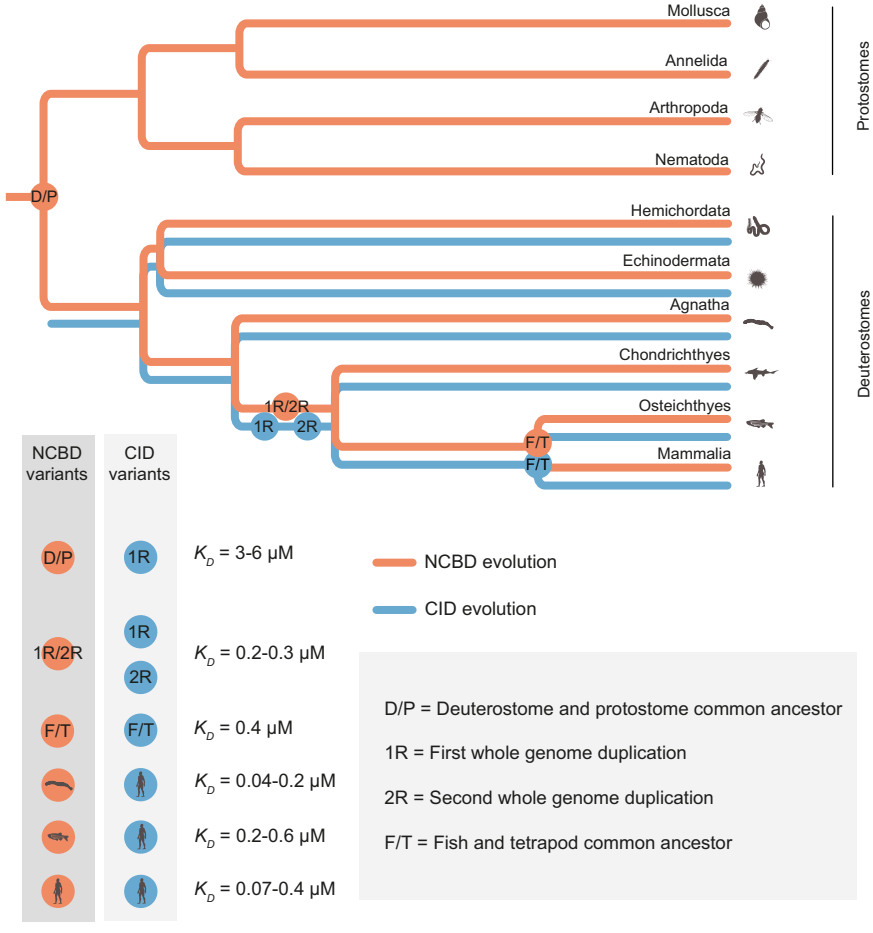

Figure 4. The evolution of NCBD and CID along with affinities of complexes from ancestral and present-day species. Previous phylogenetic analysis revealed that the NCBD domain was present in both deuterostome and protostome animals. CID, however, was only found in deuterostomes, indicating that this domain emerged in the proteome of an early deuterostome ancestor. Ancestral variants of NCBD and CID were previously reconstructed using ancestral sequence reconstruction. The affinity of the complex between the $NCBD_{D/P}$ and CID_{1R} or CID_{2R} variant was 3-6 μM. This was around 10-20-fold lower than for the $NCBD_{1R/2R}$ variant in complex with CID_{1R} or CID_{2R}, indicating that NCBD underwent some changes to increase the affinity for CID after its emergence in the proteome. The high affinity was further maintained in the fish/tetrapod ancestor proteins and in present-day zebra fish, lamprey and human.

Methodology

Biomolecular kinetics

The pioneering studies of the kinetics of chemical reactions were performed in the late 1800's. The initial experiments were performed on simple systems as opposed to the complex biomolecules that are studied today. These experiments led to the formulation of the law of mass action, which constitutes the fundament in kinetic theory and which states that the rate of a chemical reaction is proportional to the concentration of reactants.

The strength in the utilization of kinetic methods for studies of biomolecular interactions lies in the ability to resolve interaction mechanisms. In this context, the word "mechanism" implies a full understanding of the pathway of a molecule when going from state A to state B. As opposed to a kinetic method, an equilibrium-based method can never provide us with this answer. By the time equilibrium is reached, we can learn about the distribution of molecules in state A and state B, but the information regarding the path that the molecules took is lost. With a kinetic method, on the other hand, we are (in favourable cases) able to monitor formation and break-down of possible intermediates along the reaction pathway and are thus able to obtain information regarding the path between the initial and the final state. This is however not always straightforward. Binding and folding reactions with several intermediates along the reaction pathway can appear as simple two-state systems if the intermediates are only slightly populated under the experimental conditions. Furthermore, when using fluorescence detection, the lack of knowledge about the fluorescence yields of the molecular species in the reaction scheme can render elucidation of a uniquely defined model impossible. Nevertheless, information from different methods (i.e a combination of kinetic and equilibrium methods) as well as experiments performed under different conditions can in favourable cases lead to a hint about the mechanism of binding for the system under study. In the following sections I will discuss two commonly used methods in studies of biomolecular binding and folding reactions, that I have used in my thesis projects.

Stopped-flow

Stopped-flow is one of the most frequently used rapid mixing techniques in studies of biomolecular reactions (Figure 5). The technique allows for rapid

mixing of two solutions, which occur on the millisecond timescale. The method is often used to study binding reactions that occur on the milliseconds-to-seconds timescale, and the binding reaction is initiated upon mixing of the two reaction components. The method can also be used to study folding reactions by mixing with a chemical denaturant or a solution with a different pH value. The measurement cell is connected to a system for optical detection, allowing for monitoring absorbance, fluorescence emission or fluorescence polarization.

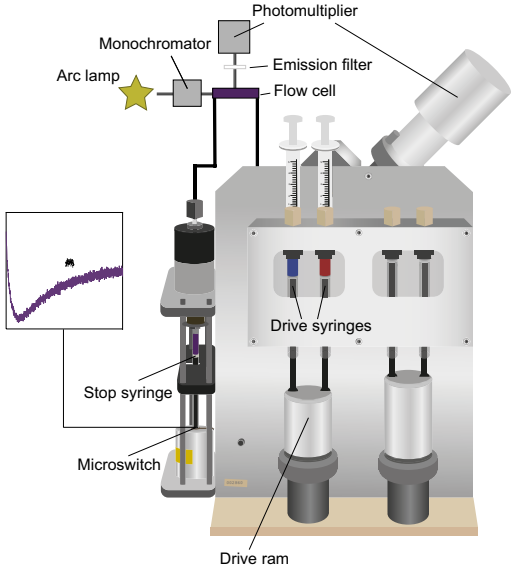

Figure 5. A semi-schematic illustration of the stopped-flow instrument. The two drive syringes are filled with solutions of the protein and ligand, respectively. When the measurement is conducted, the drive ram pushes the solutions to a flow cell, where the two solutions are mixed. This pushes the previous liquid in the flow cell towards the stop syringe, which triggers a microswitch and recording of the binding reaction starts. This figure shows the set-up for monitoring the binding reaction with fluorescence. A monochromator is used to set the excitation wavelength (280 nm when tryptophan is used as a probe). The emission of light from tryptophan often changes upon binding and this is used to monitor the reaction. The emitted light is detected after passage through an emission filter, commonly a >320 nm long-pass filter. This signal is amplified and converted into an electrical signal using a photomultiplier. The resulting kinetic trace shows up on the computer screen.

Temperature-jump

The temperature-jump method utilizes an electric discharge to rapidly increase the temperature of the solution. This perturbs the system under study and the relaxation to equilibrium can be monitored, commonly by using fluorescence detection. The increase in temperature takes place on the microsecond timescale, thus enabling studies of reactions that take place on the microseconds-

to-milliseconds timescale. The cooling of the cell back to the start temperature takes place in seconds and sets the limit for how slow reactions that can be monitored. The temperature-jump method is in general more suitable for studying folding reactions as compared to binding reactions, as the equilibrium constants for folding commonly display a higher temperature dependence. This is because the reaction enthalpy, ΔH, is generally higher for folding reactions as compared to binding reactions. However, weak binding reactions with K_D-values in the high μM range are often too fast for rapid mixing methods and could instead be measured with for example temperature-jump.

Fitting and analysis of kinetic data

When performing a kinetic experiment, there are two different kinds of information that we can obtain. One is qualitative information, such as: what is the mechanism? Or: how many reaction steps are included in this mechanism? The other is quantitative information, which are estimates of coefficients that can be extracted by fitting the data set to suitable equations. These coefficients are more or less abstract in their nature, but can in many cases be related to intuitive and logical concepts. The methods available for data fitting has developed much over the years. Fitting methods that relied on linearization of data were common practice for many years, but the development of modern computers and non-linear regression have passed these methods to history. Nowadays, there are two common practices when it comes to fitting kinetic data: 1) Fitting to derived analytical expressions and 2) Fitting the data directly to a model using numerical integration. I will discuss these two methods in detail and present pros and cons for each of these approaches.

Fitting to analytical solutions

Fitting of a kinetic data set to an analytical solution is by far the most common method for fitting kinetic data in the scientific literature at the moment. The analytical solution is an exact solution to the differential equation that is set up for any reaction mechanism based on the law of mass action. When fitting to an analytical solution, the kinetic data is often fitted in two steps. In the first step, the kinetic transient is fitted to an exponential function in order to extract rate constants (k_{obs}), amplitudes and endpoints. In the next step, these parameters are plotted against concentration of reactant (A or B). These plots provide a valuable clue regarding the quality of the obtained data, the number of apparent steps in the reaction mechanism and the plausibility of alternative mechanisms and is therefore a mandatory step in the fitting procedure regardless of which fitting method is employed in the end. The extracted data (commonly the concentration dependence of k_{obs}, although the amplitudes and endpoints contain information as well) is then fitted to an analytical function for the particular model under testing, and estimates of the microscopic rate

coefficients in the reaction scheme can be extracted. The analytical solutions are obtained by solving the differential equations with integration or matrix algebra. This is fairly straightforward (and has already been done) for the simplest binding mechanisms, but the difficulty in solving these equations increases rapidly with an increasing number of steps in the reaction mechanism. For this reason, derivation of analytical expressions for more complex reaction mechanisms often relies on simplifying assumptions about the mechanism. For example, the commonly used analytical solution for an induced fit mechanism assumes a rapid binding equilibrium such that $k_{-1} >> k_2 + k_{-2}$ in order for the equation to apply (Figure 6a)[25]. Furthermore, fitting to analytical functions relies on the assumption about pseudo-first order conditions in the experiment. Binding reactions are typically second order reactions, meaning that the rate of the reaction is dependent on the concentration of both A and B. However, in order to simplify the derivation of analytical expressions, experiments are often performed under conditions where [B] >> [A] (or vice versa), so called pseudo-first order conditions. Under these circumstances, the majority of B will not engage in a complex with A and thus the concentration of reactant B will remain essentially constant during the binding reaction. For a simple two-state system under pseudo-first order conditions, the kinetics will follow a single exponential function and k_{obs} increases linearly with concentration of B (Figure 6b). In reality, it might be difficult to perform experiments under such conditions. This is because increasing the concentration of B might cause the reaction to become too fast to monitor for the chosen kinetic method. Decreasing the concentration of A, on the other hand, might result in a poor signal-to-noise ratio.

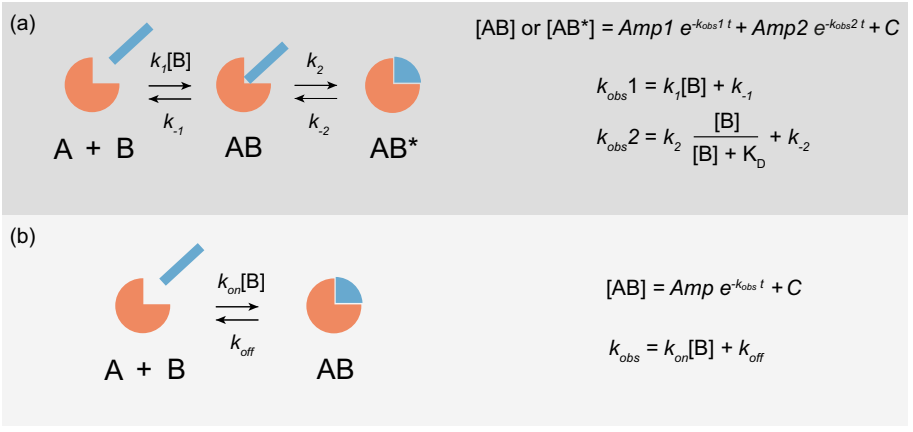

Figure 6. Analytical expressions for two different mechanistic models. (a) The analytical expressions for the two kinetic phases that may be observed in an induced fit scenario under pseudo-first order conditions and assuming a fast preequilibrium binding scenario where $k_{-1} >> k_2 + k_{-2}$. The parameters are defined as follows: Amp = kinetic amplitude of the kinetic phase, k_{obs} = observed rate constant of the kinetic phase, t = time and C = a constant describing an offset. (b) The simplest one-step mechanism for binding under pseudo-first order conditions. One kinetic phase is observed.

Fitting by numerical integration

By using the numerical integration fitting method, many of the shortcomings of fitting to analytical expressions can be overcome. This method is neither dependent on the derivation of analytical expressions, nor on assumptions about pseudo-first order conditions. Furthermore, global fitting of a full kinetic data set directly to a model using numerical integration considers all information that is stored in the kinetic traces (that is rate constants, amplitudes and endpoints). When fitting data to analytical solutions, some of this information is often omitted. The method of numerical integration is based on division of the exponential curve into linear segments at very short time intervals (Figure 7a). As the initial concentrations are known, the law of mass action gives the concentrations of reactants and products at each of these time points. As opposed to the analytical solutions which are exact, this method is obviously based on an approximation. However, given that short enough segments are used, this approximation gets close enough to the analytical expression (Figure 7b). Still, the analytical fitting method is useful in the initial fitting process, in which the number of reaction steps that can be accurately resolved are obtained. This can serve as a guide towards plausible reaction mechanisms that can further be tested with numerical integration methods.

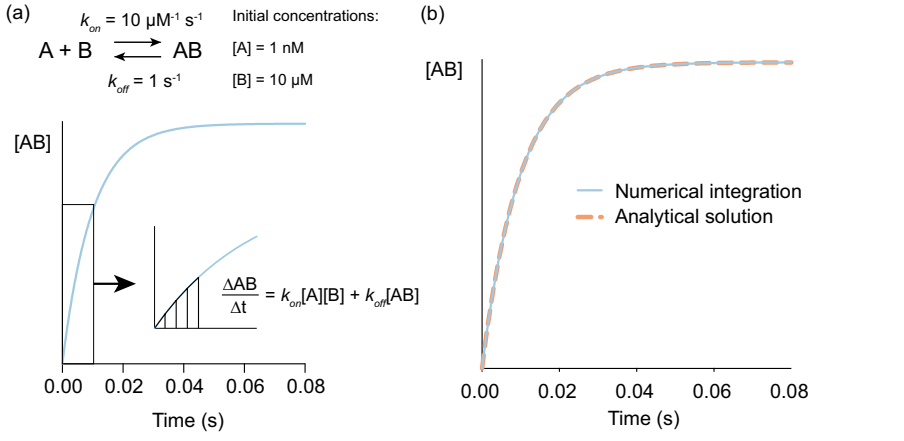

Figure 7. Fitting with numerical integration versus fitting to analytical solutions. (a) A kinetic trace that was simulated using numerical integration for a simple one-step binding reaction. When fitting with numerical integration, the kinetic curve is built up from very narrow segments. The slope of each segment is calculated using the differential equation that is given in the figure. (b) A comparison of the results from simulation with numerical integration and analytical solution, using the initial concentrations and rate constants that were defined in (a). The numerical integration method is very precise when the segments used are short enough.

Transition state analysis with ϕ-values

If we consider a simple two-state protein binding reaction, the protein can reside in two different states: the unbound state and the bound state. In between these states is the transition state, which is a high-energy state that the proteins pass when going from the unbound state to the bound state and vice versa. The transition state is an extremely short-lived state which consists of an ensemble of conformations[25]. Thus, the transition state cannot be isolated, but it is possible to study it with indirect methods such as ϕ(phi)-value analysis and linear free energy relationships. The ϕ-value analysis was initially developed for folding reactions[40], but has been applied on several protein-protein interactions involving IDPs[11].

The ϕ-value analysis and linear free energy relationships rely on the introduction of multiple conservative mutations spanning as much of the protein sequence as possible. For each mutant, the association rate constant (k_{on}) and dissociation rate constant (k_{off}) are obtained from kinetic experiments. From the comparisons between mutant and wildtype, the ΔΔG for the transition state ($\Delta\Delta G_{TS}$) and ΔΔG for the bound state ($\Delta\Delta G_{EQ}$) can be calculated for each mutant using the obtained rate constants (Equations 1-2). The ϕ-value is obtained from the ratio of these values (Equation 3). The ϕ-values typically range between 0-1, where a value of 0 indicates that the mutated residue does not form any native interactions in the transition state and a value of 1 suggests that the

mutated residue engage in fully formed native contacts in the transition state. Many ϕ-values that span an entire binding interface can lead to clues about where along the reaction coordinate native contacts are being formed. The values for $\Delta\Delta G_{TS}$ and $\Delta\Delta G_{EQ}$ can also be plotted in a linear free energy relationship plot to obtain linear free energy relationships. The slope in this plot constitutes an average ϕ-value and can thus be interpreted in the same way.

$$\Delta\Delta G_{TS} = -RT \ln \frac{k_{on\,MT}}{k_{on\,WT}} \qquad \text{(Equation 1)}$$

$$\Delta\Delta G_{EQ} = -RT \ln \frac{K_{D\,WT}}{K_{D\,MT}} \qquad \text{(Equation 2)}$$

$$\phi = \frac{\Delta\Delta G_{TS}}{\Delta\Delta G_{EQ}} \qquad \text{(Equation 3)}$$

$\Delta\Delta G_{TS}$ = Difference between wildtype and mutant in the free energy of the transition state (J mol^{-1})
$\Delta\Delta G_{EQ}$ = Difference between wildtype and mutant in the free energy of the bound state (J mol^{-1})
R = Molar gas constant (J K^{-1} mol^{-1})
T = Temperature (K)
k_{on} = association rate constant (M^{-1} s^{-1})
k_{off} = dissociation rate constant (s^{-1})
K_D = Equilibrium dissociation constant (M)

Other biophysical techniques that were employed in the work

Circular dichroism

Circular dichroism (CD) is commonly used as a crude estimate of the secondary structure content of proteins. The CD instrument measures the difference in absorption of left-handed and right-handed circularly polarized light and converts it to ellipticity (θ) which has the unit millidegrees (mdeg). The CD spectrum between wavelengths 190-260 nm display characteristic patterns depending on the secondary structure content. The CD spectrum for an alpha helical protein, for instance, displays two minima in the spectrum at 208 and 222 nm (Figure 8a). Often, the CD signal is used as a probe for structural content in denaturation experiments. For example, when monitoring unfolding

of an alpha helical protein, it is common to monitor the CD signal at 222 nm over a series of increasing denaturant or heat. The loss of signal at 222 nm can be used to monitor unfolding of the protein (Figure 8b) and the unfolding curve can be fitted to obtain parameters related to the folding stability of a protein (Figure 8c).

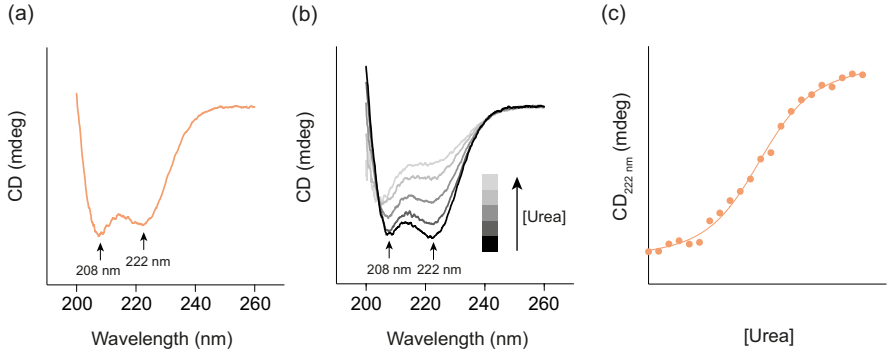

Figure 8. CD spectrum and unfolding of an alpha-helical protein. (a) A CD spectrum for a typical alpha-helical protein. The CD spectrum displays two minima at 208 and 222 nm. (b) The CD signal at 222 nm disappears as the protein unfolds at higher concentrations of the denaturant urea. (c) The CD signal at 222 nm plotted against urea concentration. By fitting to a two-state equation, parameters related to the folding stability of the protein can be estimated.

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is used to obtain information regarding the affinity, the enthalpic and entropic contributions to affinity and the stoichiometry of biomolecular interactions. The instrument has two insulated cells: one measurement cell and one reference cell which contains a solution with the same heat capacity as that in the measurement cell. The experiment is conducted by titrating ligand from a syringe into the measurement cell which contains a solution of the interacting protein. Upon binding, heat is either taken up from or released to the solution in the measurement cell, due to breaking and formation of weak interactions in the binding reaction. Under constant pressure conditions, this heat corresponds to the enthalpy (ΔH) of the binding reaction. The instrument uses power to keep the temperature of the reference cell equal to that of the measurement cell, hence the name isothermal (iso = equal). The power peaks are integrated, normalized with respect to amount of injected ligand and plotted against molar ratio ([ligand]/[protein]). The data is fitted to a suitable equation describing the system under study, to obtain estimates of affinity (K_A), enthalpy and stoichiometry[41]. The contribution from entropy (ΔS) can be calculated using Equation 4 and 5.

$$\Delta G = -RT \ln K_A \qquad \text{(Equation 4)}$$

$$\Delta G = \Delta H - T\Delta S \qquad \text{(Equation 5)}$$

R = Molar gas constant (J K^{-1} mol^{-1})
T = Temperature (K)
K_A = Equilibrium association constant (M^{-1})
ΔS = Entropy difference between the bound and unbound state (J K^{-1} mol^{-1})
ΔH = Enthalpy difference between the bound and unbound state (J mol^{-1})

Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) was employed to obtain structural information about the ancestral NCBD and NCBD-CID complex in solution. The NMR technique utilizes the magnetic properties of nuclei with an uneven number of protons and/or neutrons. The most commonly used nuclei in protein-NMR are ^{1}H, which is highly abundant in nature, and ^{15}N and ^{13}C, which needs to be enriched in the sample.

Magnetic nuclei can reside in two separate spin states. The energy difference between the two spin states will depend on an intrinsic property of each nucleus and on the strength of the magnetic field. This energy difference between the spin states is very small, and corresponds to the frequency of a radio pulse.

The sample experiences a strong constant magnetic field which is applied externally. However, each nuclei experiences a difference in the electronic environment that leads to small local differences in the magnetic field strength. Thus, the frequency of the radio pulse corresponding to the energy gap will be different for each nucleus in the sample. This frequency is normalized towards a reference frequency and multiplied with a million, resulting in the chemical shift with unit parts per million (ppm).

Using combinations of different NMR spectra, one can assign chemical shifts to most backbone and sidechain nuclei in the protein. This information can be used to find nuclei that are close in space by recording so called nuclear Overhauser effect (NOE) experiments. The nearness in space between two nuclei in such an experiment will be proportional to the intensity of the NOE crosspeak. These peaks and their intensities, can further be used as restraints in computer programs to find a model for the protein structure.

Present investigations

Background and aim(s) of the thesis projects

As reviewed in the previous sections, IDPs are involved in fundamental cellular processes and have likely acted as drivers in the development of multicellular organisms. Since their discovery, much research efforts have been focused around understanding how the biophysical properties of IDPs translates to their function. Despite the research boom on IDPs which has led to major discoveries about the biophysical characteristics that underlie their interactions, the questions regarding how such interactions evolve on a molecular level still remain unresolved. In the first part of my thesis projects, I have addressed such questions. Using previously reconstructed ancestral variants of the NCBD and CID, the aim was to characterize these proteins with respect to their structure, stability and transition state. In the second part of my thesis projects, the aim was to investigate the impact of disordered regions flanking the binding interface of the human CID protein.

The work in this thesis have contributed to knowledge regarding how evolution shapes protein-protein interactions involving IDPs as well as the role of disordered regions flanking the binding interfaces in protein-protein interactions.

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

The established protein-protein interactions in the interactome have evolved to maintain proper function of the organism. NCBD from CREBBP/p300 interacts with multiple ligands, and it is intriguing to understand how such a multiligand-binding protein adapts to bind new ligands. Previously, a phylogenetic analysis was carried out for NCBD and one of its interaction partners CID. The analysis showed that NCBD was present in the common ancestor of the deuterostome and protostome clades in the animal kingdom, whereas CID likely emerged in the proteome after the divergence of these lineages. The initial resurrection of ancestral NCBD and CID variants showed that the high-affinity interaction that is observed among present-day vertebrates likely emerged from an early NCBD-CID complex with low affinity (see Figure 4 in Introduction)[35].

In paper I, the aim was to dissect the structural and thermodynamic basis for the evolutionary adaptation of the NCBD-CID interaction. Two ancestral NCBD-CID complexes were analysed. The "Cambrian-like" complex consisted of NCBD from the common ancestor of deuterostomes and protostomes ($NCBD_{D/P}$) in complex with CID from the time of the first whole genome duplication (CID_{1R}). The "Ordovician-Silurian" complex consisted of NCBD from the time of the first and second whole genome duplication ($NCBD_{1R/2R}$) in complex with CID_{1R}. NCBD from human CREBBP and CID from human NCOA3 was used as a comparison to the ancestral protein complexes (Figure 9a).

First, the structures of the ancestral and human NCBD-CID complexes were solved by NMR. The NMR structures revealed small structural differences between the three complexes (Figure 9b-c). The orientation of the helices in NCBD is slightly different between all three variants, and the third helix of CID is more well-defined and forms more intermolecular contacts with NCBD in the Ordovician-Silurian and human complex as compared to the Cambrian-like complex. The complexes also differ in compactness. The Cambrian-like complex displays the highest surface-accessible surface area (SASA) compared to the other two complexes.

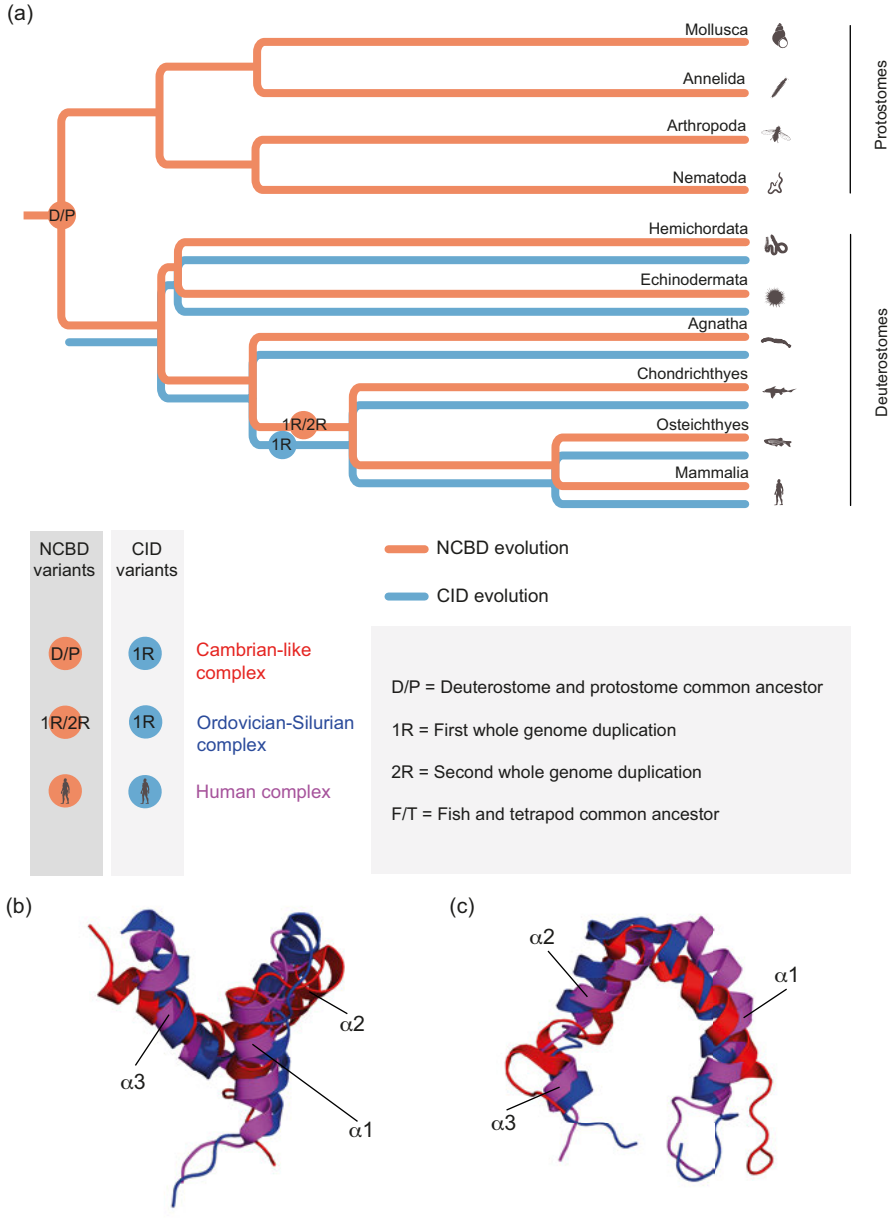

Figure 9. NMR structures of ancestral and human NCBD-CID complexes. (a) A phylogenetic tree with the different ancestral NCBD and CID variants pointed out. (b) The NMR structures for the Cambrian-like, Ordovician-Silurian and human complex were overlayed. Here, only the NCBD proteins are shown for clarity. The orientation of the helices of NCBD is slightly different between all three complexes. (c) An overlay of the three complexes where the NCBD proteins have been omitted for clarity. Helix 1 and 2 display different orientations. The third helix of CID is most well-defined in the Ordovician-Silurian and human complex.

My contribution to this study consisted of examining the thermodynamic basis for this interaction with isothermal titration calorimetry (ITC; Figure 10a). The data were collected over a broad pH range (pH 5.2-8.2) and the dependency of the affinity on pH for all complexes displayed broad transitions, indicating that many titratable groups are involved in modulating the affinity (Figure 10b). The data show similar pH-dependencies for the Cambrian-like and Ordovician-Silurian complex, with the affinity increasing for both complexes at lower pH values and thus maintaining around a 5-10-fold difference in affinity at all pH-values. The pH-dependence of the human complex goes in the opposite direction, and the affinity for the human and Cambrian-like complex converges at the lower pH values.

The enthalpic and entropic contributions to affinity were further analysed (Figure 10c). At all pH values, the complexes displayed favourable enthalpic contributions and unfavourable entropic contributions to the free energy. Enthalpy-entropy compensation is observed in the data set and no clear trend regarding the pH-dependency of these parameters could be discerned. While this could reflect co-dependency of these parameters in the measurements, these data are also consistent with a protein-protein complex with large malleable interaction surfaces, which can rearrange slightly upon a change in conditions such as pH.

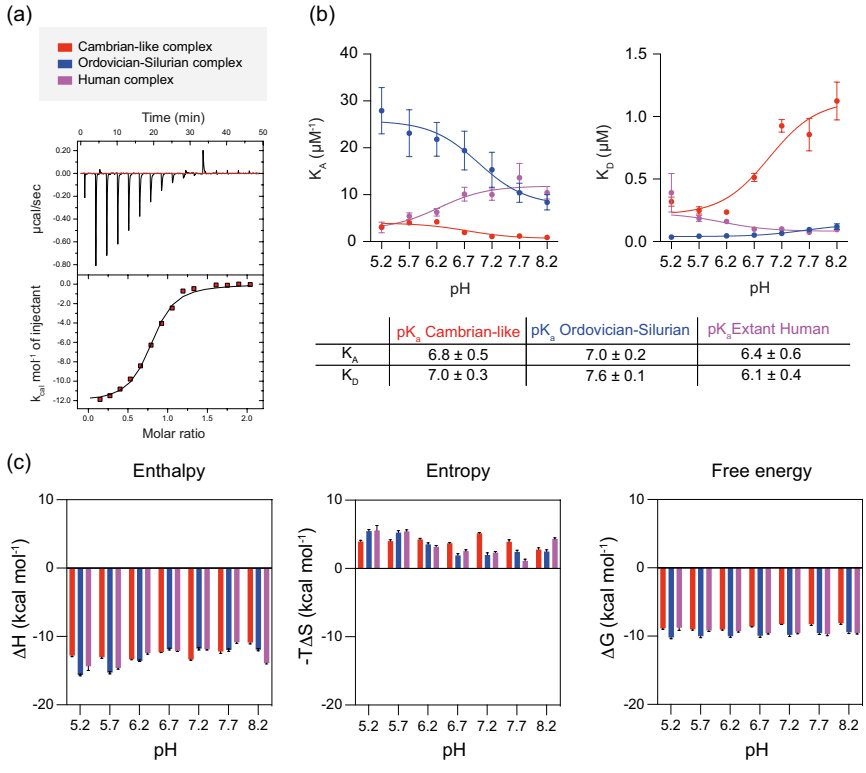

Figure 10. Dependence on pH for thermodynamic parameters of the ancestral and human NCBD-CID complexes. (a) An example ITC thermogram for the Cambrian-like complex. The data is fitted to a two-state equation to obtain estimates of affinity, stoichiometry and thermodynamic parameters. (b) The pH-dependencies of K_A and K_D for the ancestral and human complexes. The data were fitted to a two-state equation and the apparent pK_a values for each interaction were extracted. (c) Enthalpic and entropic contributions to the free energy at each pH value.

Paper II: Mapping the transition state for a binding reaction between ancient intrinsically disordered proteins

In order to further characterize the oldest Cambrian-like complex, the transition state of this complex was probed using ϕ-value analysis. Previously, the transition state of the human NCBD-CID complex was probed using ϕ-value analysis and linear free energy relationships. This showed that the majority of the packing of hydrophobic contacts occurred in a cooperative manner after the rate-limiting activation barrier[42]. In order to probe the transition state for the Cambrian-like complex, several conservative mutations and a few non-conservative mutations were introduced in both $NCBD_{D/P}$ and CID_{1R}, which targeted native contacts in the binding interface of the proteins. Unfortunately, many of the mutations that targeted native hydrophobic contacts resulted in great destabilization of the Cambrian-like NCBD-CID complex and made

kinetic analysis impossible due to high k_{obs} values and low kinetic amplitudes. Nevertheless, a few ϕ-values were obtained for this complex and these were compared to the previously obtained ϕ-values for corresponding positions in the human complex (Figure 11a-b)[42].

In general, the ϕ-values are higher for the Cambrian-like complex, in particular for the positions in NCBD, indicating a transition state with more native interactions in the Cambrian-like complex as compared to the human complex. A few positions, such as L2087 and L2096 displayed large differences in their ϕ-values, indicating a local rearrangement in the transition state structure around these positions.

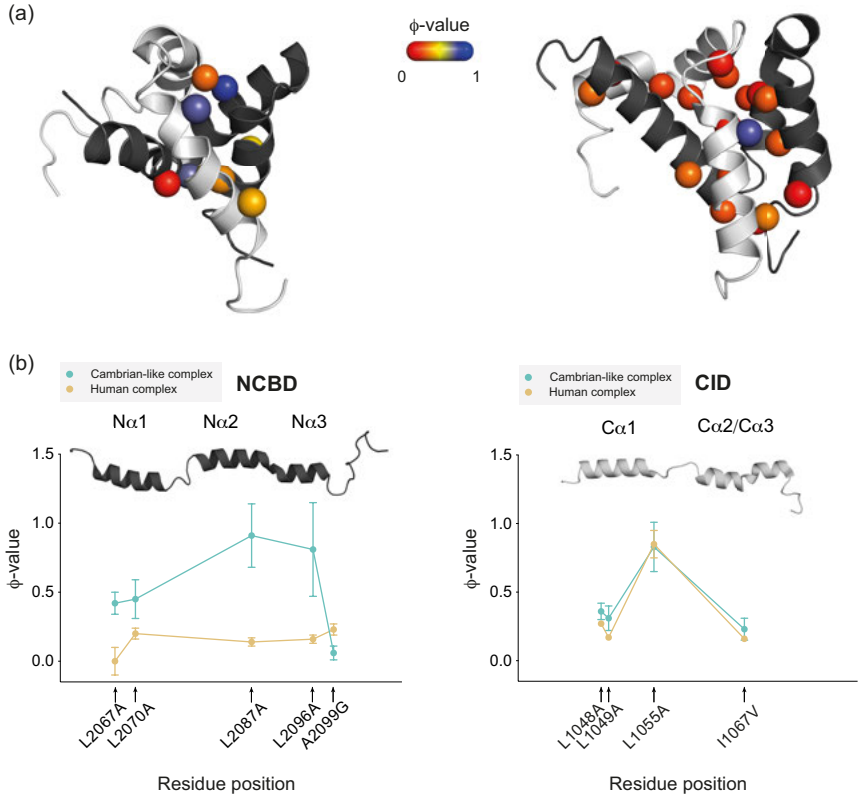

Figure 11. Transition state analysis for the Cambrian-like and human complex. (a) The resulting ϕ-values from conservative mutations in the Cambrian-like protein complex (left) and present-day human complex (right). NCBD is coloured dark grey and CID in light grey. The ϕ-values for the human complex were from a previous study[42]. The structure of the Cambrian-like complex was modelled using PDB entry 6ES5[43] and the human complex using PDB entry 1KBH[44]. (b) A comparison between the ϕ-values for the Cambrian-like (teal) and human (gold) complex at the corresponding positions in the two complexes. The ϕ-values for NCBD are shown to the left and ϕ-values for CID to the right. The figure was adapted from[45].

Furthermore, helix-modulating mutations were introduced in CID_{1R} and in human CID, in order to probe for helical content in the transition state of both the Cambrian-like and human complexes. Previously, ϕ-values for the first helix (CIDα1) in human CID were obtained, and these are used here for comparison to the Cambrian-like complex[36]. The overall mechanism of helix formation appears similar for the Cambrian-like and human complex, as judged by the similar trends in the ϕ-values for the helix-modulating mutations (Figure 12a). Linear free energy relationships suggest, however, that CIDα1 is slightly more populated in the transition state of the human complex as compared to the ancestral complex (Figure 12b). Thus, our data show large similarities in helix formation in the transition state for the ancestral and human complex. The higher degree of helix formation in the transition state for human CIDα1 could reflect the higher helical propensity for this helix[35].

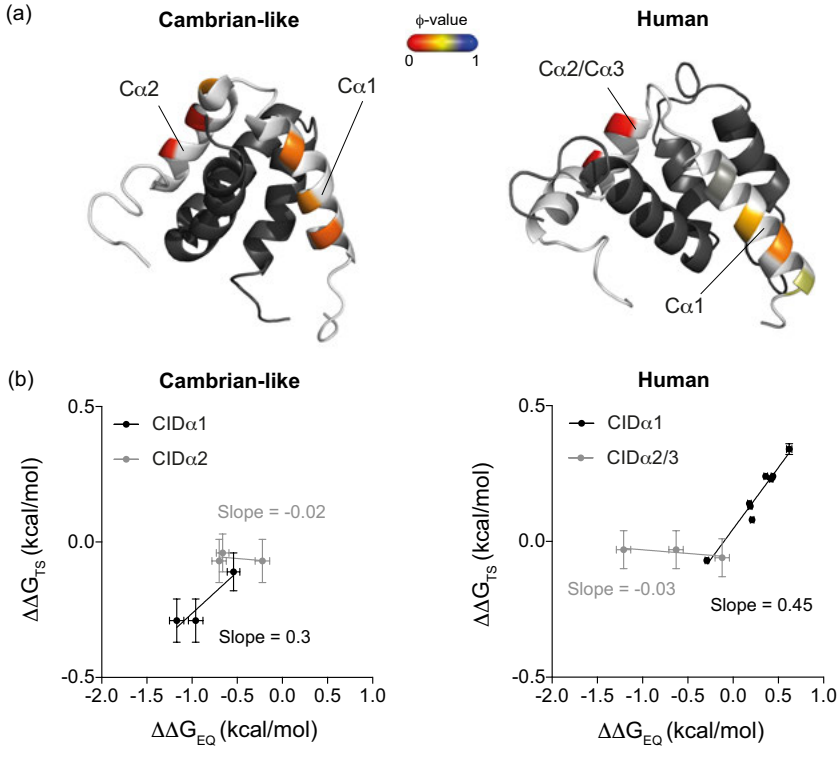

Figure 12. Transition state analysis for helix formation in the Cambrian-like and human complex. (a) The ϕ-values for helix-disrupting mutations mapped onto the Cambrian-like structure (left) and the ϕ-values for helix formation of CID in the present-day human complex (right). NCBD in dark grey and CID in light grey. The ϕ-values for the first α-helix were obtained previously[36]. (b) Linear free energy relationships for helix formation of the Cambrian-like complex (left) and present-day human complex (right). The slopes have been calculated for α-helix 1 (CIDα1) and α-helix 2 and 3 (CIDα2 or CIDα2/3) separately. The figure was adapted from[45].

Paper III: High affinity between CREBBP/p300 and NCOA evolved in vertebrates

Previous phylogenetic analysis of the NCBD-CID interaction has shown that NCBD was present in the common ancestor to the deuterostome and protostome lineages, whereas CID likely emerged after the divergence of these lineages. Subsequent biophysical analysis showed that by the time of the two whole genome duplications that occurred in an ancestor of modern vertebrates, a high-affinity complex had been established that was further maintained in several vertebrate species[35]. In paper III, we sought to analyse the affinity of the NCBD-CID complex in a broader range of deuterostome animals. NCBD and CID variants from human (*H. sapiens*), zebra fish (*D. rerio*), sea lamprey (*P. marinus*) and sea urchin (*S. purpuratus*) were analysed with respect to their binding affinity using ITC. Two variants of CID were included from the sea lamprey lineage, denoted NCOAa and NCOAb.

The affinity of the NCBD-CID interaction was highly similar in human, zebrafish and sea lamprey, with K_D-values ranging between 0.2-0.6 μM. However, the affinity of the sea urchin NCBD-CID complex was strongly reduced and displayed an average K_D-value of 15 μM (Figure 13a). NCBD from sea urchin was tested with CID from human, an vice versa, resulting in low affinities in both cases (K_D-values of 12 and 4 μM, respectively). This showed that amino acid alterations in both the NCBD and CID sequence in sea urchin contributed to the lower affinity. A comparison of the sequences of NCBD and CID from sea urchin and human disclose several substitutions that might partially explain the large difference in affinity between the proteins from these species. Particularly interesting are the R2072Q, Q2084K, K2107H and Y2108Q substitutions in NCBD (comparing sea urchin to human NCBD; Figure 13b). In CID, the substitutions in the LXXLL motif (L1052F and L1055F, comparing sea urchin to human CID) as well as the R1069N (comparing sea urchin to human CID) substitution in the C-terminal region are particularly notable (Figure 13b). Previous data showed that an R1069A mutation in CID_{1R} caused a 3-fold decrease in affinity of the ancestral Cambrian-like NCBD-CID complex[45]. However, it is likely that many residue substitutions act together to modulate the affinity of this protein complex.

Altogether, the data in paper III shows that the high-affinity interaction between NCBD and CID likely emerged in the vertebrate lineage.

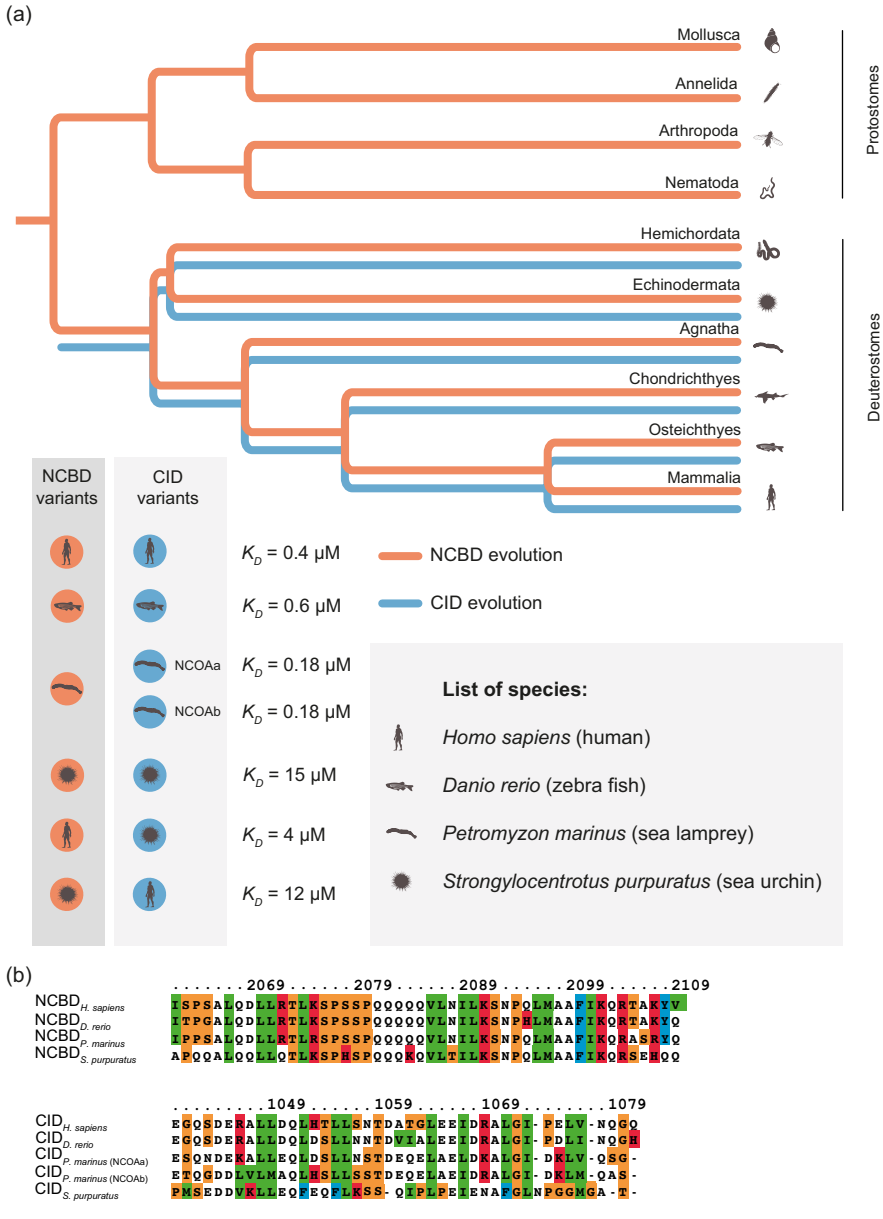

Figure 13. Affinities of NCBD-CID complexes from different deuterostome species. (a) The affinities were estimated using ITC. Three vertebrates (human, zebra fish and sea lamprey) and one non-vertebrate (sea urchin) were included in the analysis. The affinity of the NCBD-CID interaction in the vertebrate species were all in the same order of magnitude, with K_D values ranging from 0.2-0.6 μM. The affinity of the NCBD-CID interaction in the sea urchin was 20-50-fold lower, with an average K_D of 15 μM. (b) Alignments of the amino acid sequences for the NCBD and CID variants included in the analysis.

Paper IV: The dynamic properties of a nuclear co-activator binding domain are evolutionarily conserved

In paper IV, the dynamic properties of NCBD were investigated from an evolutionary perspective. NCBD constitute an interesting model system, due to its multiligand binding capacity and molten globule-like properties. In order to examine how evolution shaped the biophysical properties of the NCBD protein, we utilized a combination of NMR spectroscopy, CD spectroscopy and kinetics to study the ancestral NCBD protein from the common ancestor of deuterostomes and protostomes ($NCBD_{D/P}$) in detail. The structure of the $NCBD_{D/P}$ variant was solved by NMR and compared to a previous NMR structure of human NCBD[46]. The NMR structure showed that the overall fold is similar for the ancestral and human NCBD, but minor differences are present. The main differences are a change in the orientation of the first and third helix (Figure 14a). In order to evaluate the conclusions with respect to the uncertainty in the prediction of the ancestral $NCBD_{D/P}$ sequence, an "AltAll" $NCBD_{D/P}$ variant ($NCBD_{D/P}^{AltAll}$) was included in the analysis. The $NCBD_{D/P}^{AltAll}$ consist of a combination of the second most probable amino acid at every position where the probability score is lower than ≈0.9, and can be used to test for the robustness of the results obtained with reconstructed ancestral variants[47]. The structure of the $NCBD_{D/P}^{AltAll}$ variant was solved with NMR spectroscopy and compared to the structures of $NCBD_{D/P}$ and human NCBD. The $NCBD_{D/P}^{AltAll}$ display a similar topology to $NCBD_{D/P}$ and human NCBD. There are differences between $NCBD_{D/P}$ and the $NCBD_{D/P}^{AltAll}$ variant mainly in the orientation of the first and third helix. (Figure 14b). These results show that the overall fold of NCBD has likely been preserved throughout evolution. The small differences that are observed between the $NCBD_{D/P}$ and human NCBD are observed also for the $NCBD_{D/P}$ and $NCBD_{D/P}^{AltAll}$, and thus likely reflect the variation of folds that can be captured within the uncertainty of the resurrected $NCBD_{D/P}$ sequence.

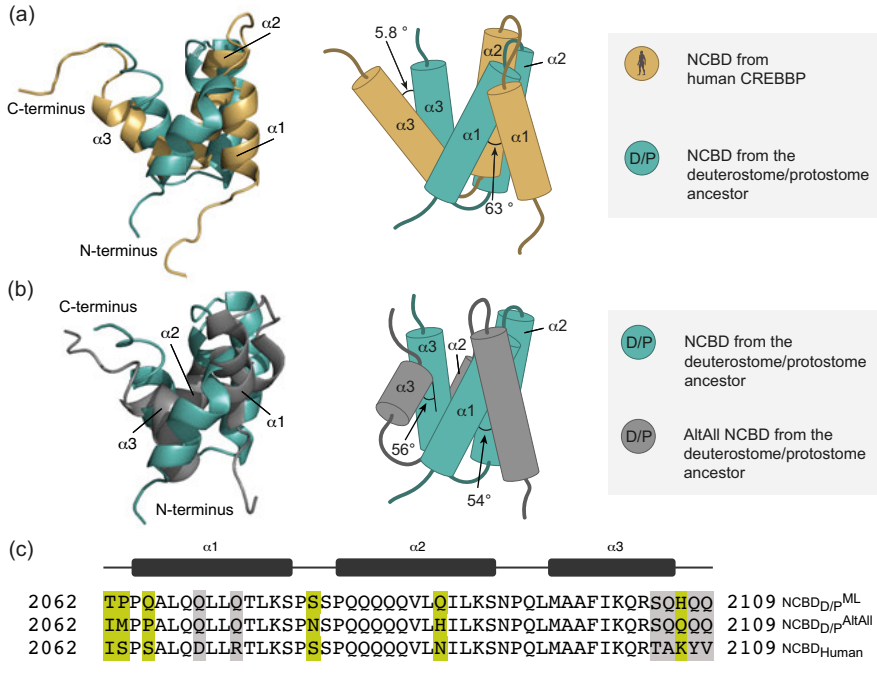

Figure 14. Structural differences between ancestral and human NCBD. (a) An overlay of the solution NMR structure of $NCBD_{D/P}$ (teal) and the previously solved structure for human NCBD from CREBBP (gold)[46]. (b) An overlay of the NMR structures of $NCBD_{D/P}$ (teal) and $NCBD_{D/P}^{AltAll}$ (grey). (c) The amino acid sequences for the different NCBD variants. The positions which differ between $NCBD_{D/P}$ and $NCBD_{D/P}^{AltAll}$ are highlighted in lime green and additional positions that differ between human NCBD and $NCBD_{D/P}$ are highlighted in grey.

Furthermore, the folding stability of the NCBD variants were investigated using CD-monitored urea and heat denaturation. Temperature denaturation experiment for human NCBD showed a non-cooperative transition and a relatively high degree of residual alpha-helical structure at higher temperatures (363 K), typical for IDPs (Figure 15a). The temperature denaturation curves look highly similar for $NCBD_{D/P}$ and $NCBD_{D/P}^{AltAll}$, showing that the dynamic properties of NCBD have likely been preserved during evolution (Figure 15b). The folding free energy could not be quantified for the wildtype $NCBD_{D/P}$ protein, due to the low stability and lack of a well-defined native baseline in the urea denaturation experiments. However, upon substitution of a Thr with a Trp residue at position 2073, the stability increased for all variants and the denaturation curve could be fitted properly (Figure 15c). These experiments showed that the thermodynamic stability of the human protein has increased compared to the ancestral $NCBD_{D/P}$ and $NCBD_{D/P}^{AltAll}$. The stability of the $NCBD_{D/P}^{AltAll}$ variant was highly similar to the $NCBD_{D/P}$ (Figure 15d).
Altogether, our data show that although evolution has altered affinity for ligands and global stability, the overall fold and molten-globule-like properties

of NCBD have been preserved during evolution. This preservation highlights the importance of dynamic properties for the biological function of NCBD.

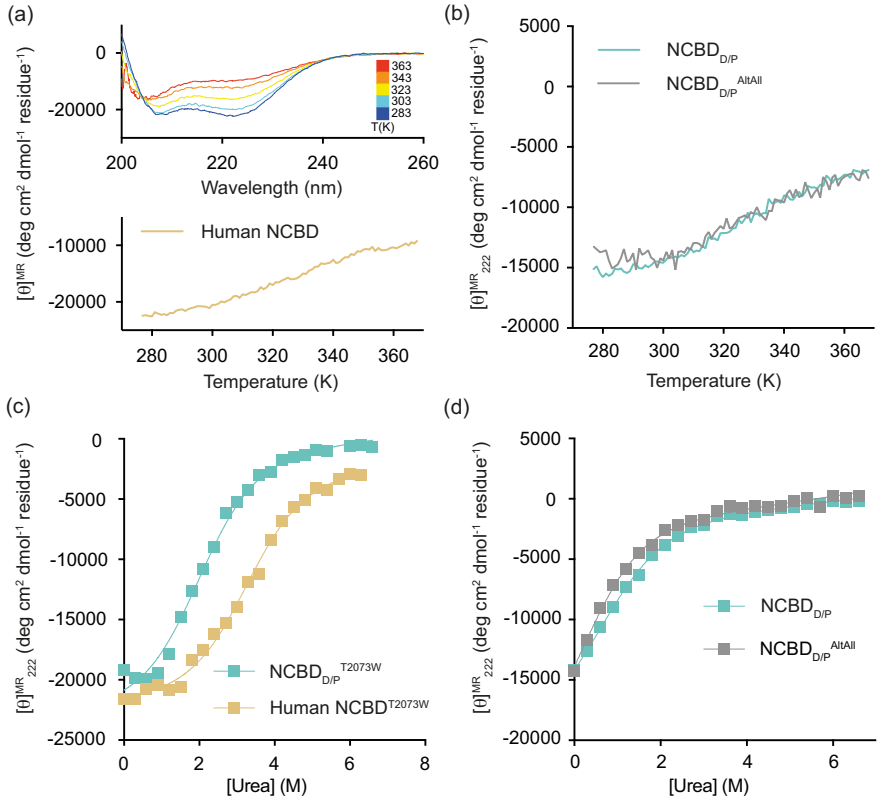

Figure 15. Heat and chemical denaturation of ancestral and human NCBD variants, monitored with CD spectroscopy. (a) Temperature denaturation for human NCBD. The spectra were taken at different temperatures (top) and the CD signal at 222 nm was used to monitor the loss of helical content (bottom). (b) Heat denaturation for $NCBD_{D/P}$ (teal) and $NCBD_{D/P}^{AltAll}$ (grey), monitored with the CD signal at 222 nm. (c) Urea denaturation of $NCBD_{D/P}^{T2073W}$ and human $NCBD^{T2073W}$. The data were fitted to a two-state sigmoidal equation to obtain the folding parameters. (d) Urea denaturation for $NCBD_{D/P}$ and $NCBD_{D/P}^{AltAll}$. The fit to a sigmoidal equation is shown, where certain parameters were shared among the data set to enable fitting despite the lack of a well-defined native baseline.

Paper V: Disordered regions flanking the binding interface modulate the affinity between CBP and NCOA

Protein domains in large proteins are often flanked by longer disordered regions. NCBD and CID are both located in disordered regions of the CREBBP/p300 and NCOA proteins, respectively. When experiments involving small protein domains are performed, the disordered flanking regions are most commonly omitted and the experiments are conducted using the minimal interaction domains. In paper V, we sought to test the impact of the disordered flanking regions and used NCBD from human CREBBP/CBP and CID from human NCOA3 as a model system for the analysis (Figure 15a). Various NCOA3 constructs were designed and tested for binding against NCBD using stopped-flow fluorimetry (Figure 15b). The association rate constant (k_{on}) and dissociation rate constant (k_{off}) were obtained for all constructs and the affinity was calculated as $K_D = k_{off}/k_{on}$. (Figure 15c). The minimal CID protein displayed the lowest affinity and a reduced k_{on} and increased k_{off} as compared to the N-CID-C variant. Including the N-terminal and C-terminal disordered regions flanking the minimal CID increased the affinity for NCBD roughly by 3-fold. Residues in both the N-terminal and C-terminal flanking regions contributed to the higher affinity and the affinity results for designed "swapped" variants (C-CID-N, C-CID and CID-N) showed that the impact on affinity was not sequence-specific. Furthermore, we quantified the rate constants and affinity over a broad range of ionic strength conditions and observed a highly similar ionic strength dependence for all NCOA3 variants, indicating no contribution from electrostatic interactions in the flanking regions on binding. Point mutations of hydrophobic residues in the N-terminal flanking region led to minor decreases in the affinity as compared to N-CID-C. This led us to hypothesize that the effect on affinity from the flanking regions is mediated by many short-lived hydrophobic and/or polar interactions. Altogether, these results demonstrate how disordered flanking regions can affect the results that are obtained in biophysical studies for protein-protein interactions. The biological significance of the results is unclear, but potentially flanking regions could function as a buffer for maintaining affinity so that mutations that drive novel functions are better tolerated.

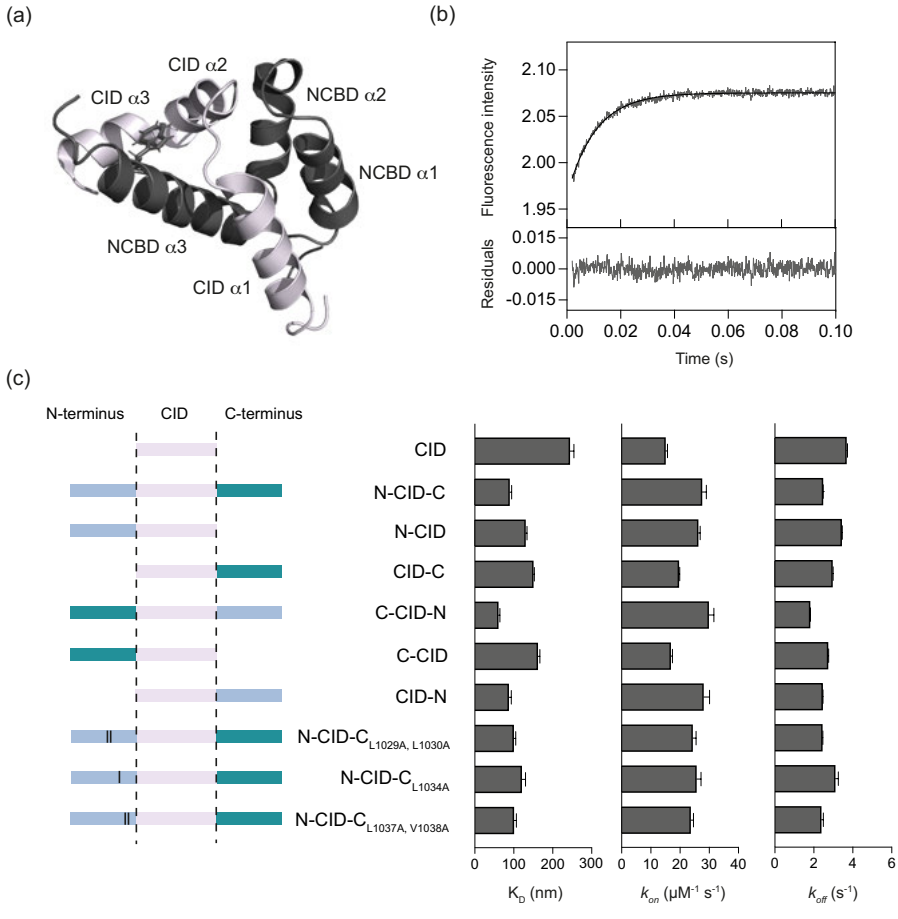

Figure 16. The impact of disordered regions flanking the CREBBP-binding interface of human NCOA3. (a) NCBD from human CREBBP (black) and different CID from NCOA3 (white) constituted the model system. The engineered Trp2108 residue (shown as a stick model) was used as a probe to monitor the binding reaction with stopped-flow fluorimetry. The structure was created in PyMol using the PDB entry 1KBH[8]. (b) A kinetic trace example from the stopped-flow measurements. The trace was fitted to a single exponential function and the residuals are shown below. (c) NCOA3 constructs of different length were tested for binding against human NCBD. The bar graphs show the results from the stopped-flow measurements for each NCOA3 variant.

Concluding remarks and future perspectives

The work done in this thesis have contributed to an increased understanding for how proteins adapt to make up a novel interaction.

The characterisation of the ancestral NCBD-CID complexes and comparison to the present-day human complex revealed large similarities in terms of overall fold and biophysical properties, despite a rather low sequence identity for the proteins (77 and 64 % sequence identity comparing $NCBD_{D/P}$ and human NCBD and comparing CID_{1R} and human CID, respectively). However, our data show that the Cambrian-like complex is less compact, has a lower helical content and fewer intermolecular contacts as compared to the human complex, which provides a biophysical explanation for the lower affinity of the Cambrian-like complex as compared to the present-day human complex.

Furthermore, we have compared the transition state and overall binding mechanism between the Cambrian-like and human NCBD-CID complexes, using ϕ-value analysis. Our results showed that the ϕ-values were overall higher for the Cambrian-like complex, which indicate a more native-like transition state for this complex as compared to the human complex. The differences in the ϕ-values are especially large for the mutations that were introduced in NCBD, which could potentially reflect ground-state effects of the introduced mutations. The $NCBD_{D/P}$ variant could be more sensitive to ground-state effects than the human NCBD due to its lower stability. It may also reflect a low robustness of the binding pathway with respect to substitutions in the NCBD and CID sequences. We have previously tested for the robustness of the binding pathway of the human NCBD-CID complex by comparing ϕ-values for different positions in the wildtype background as compared to in the background of an additional mutation in either NCBD or CID. This analysis showed that the ϕ-values for the NCBD-CID complex are not robust to changes in the sequence, and indicates a malleable binding pathway which might be a common feature among synergistic binding and folding interactions[48]. Nevertheless, the ϕ-values were overall rather similar for the Cambrian-like and human complexes, indicating an overall well-conserved binding pathway for the NCBD-CID complex. The observed differences in the ϕ-values might reflect malleability of the binding pathway of the NCBD-CID interaction which might facilitate adaptation to new binding partners.

Our analysis of the folding stability of ancestral and human NCBD showed that the $NCBD_{D/P}$ variant displays a significantly lower folding stability as

compared to the human NCBD, which likely affects the affinity for CID. Residual structure content has been shown to correlate well with affinity for some IDP interactions[36,49]. Does altering protein stability present a way for biological systems to modulate the affinity of IDPs for ligands? Based on the binding affinity data for the $NCBD_{D/P}^{AltAll}$ variant, which displays highly similar stability as the $NCBD_{D/P}$ variant but a 3-fold lower affinity for CID_{1R}, we know that some substitutions have de-correlated effects on stability and affinity. It would be intriguing to investigate the correlation between stability and affinity among a larger set of NCBD substitutions, in order to map out the correlation with stability and affinity at each residue position. Furthermore, previous binding studies on ancestral NCBD variants showed that the affinity for other ligands were kept fairly constant at each evolutionary node, while only the affinity for CID increased during this evolutionary interval[35]. Finding such specificity-altering substitutions and investigating their correlation to other parameters such as co-evolution between different ligands at the residue level, conservation, site-specific evolutionary rate or position in the protein structure would also pose an interesting project.

In my final thesis project, I have investigated the impact on affinity from disordered regions flanking the CID binding interface. Our data shows a clear impact on affinity mediated by these regions. The effect on affinity was not sequence-specific and not mediated by electrostatic forces. We therefore concluded that the contribution to affinity from the disordered flanking regions is likely mediated by non-specific short-lived hydrophobic and/or polar interactions. These results highlight the importance of trying to include larger parts of the proteins when studying their interactions biophysically. The main part of the biophysical data on proteins has been generated using isolated domains cut out from their larger protein context. This creates a bias in the literature, which prevents a full understanding of interdomain contacts and influence on binding or folding of disordered flanking regions or neighbouring globular domains in the larger protein structure. It can be extremely challenging to express and purify large full-length proteins for *in vitro* studies, as long disordered regions are especially prone to degradation and precipitation during expression and purification, and likely not all proteins are suited for this type of experimental set-up. It is often the development and improvement of new techniques that drive paradigm changes in science. The development and improvement of methods that allow for studying proteins directly in the cell environment would greatly enhance our understanding of how proteins behave in their full-length context.

Finally, it would be a shame to conclude this thesis work without mentioning the scientific progress that has been made recently using deep-learning networks, in particular when applied on the protein folding problem. These deep-learning networks can now predict protein structures with high confidence for most amino acid sequences, using large amounts of sequence data and a library of previously solved protein structures as the input information[50].

This demonstrates the utility of sequence data in the protein chemistry field, and how it can be used to provide an understanding of the evolution, structure and function of proteins and how these parameters are correlated.

Sammanfattning på svenska – oordnade proteiners evolution och interaktioner

Inuti de celler som bygger upp en människokropp utförs den största delen av arbetssysslorna av proteiner. Proteiner bistår till exempel i kemiska reaktioner som gör att vi kan tillgodogöra oss energi från mat och de sköter celldelningen som gör att våra vävnader kan förnya sig så att vi kan läka sår och åldras. Proteinerna är små molekyler som såklart är utan medvetande - ändå lyckas de lösa dessa komplicerade uppgifter åt oss.

Instruktionerna för vad som ska utföras i cellerna kommer från signaler utanför cellerna. Denna information förs vidare inom cellen med hjälp av nätverk av proteiner som interagerar med varandra, vilka dirigerar cellen till att utföra de åtgärder som behövs. De proteiner som är involverade i den här typen av reglering är ofta så kallade oordnade proteiner. De oordnade proteinerna saknar en kompakt struktur, är flexibla och kan ibland anta olika strukturer beroende på vilken miljö och sammanhang som de befinner sig i. Deras flexibilitet tros vara en fördel i deras arbete med att sköta reglering inuti cellerna – den gör att de enklare kan modifieras på olika sätt och på så sätt anpassa sig till de rådande förutsättningarna.

I min avhandling har jag studerat interaktionen mellan CREBBP och NCOA, vilkas funktion är att reglera genuttryck som respons på signaler från olika hormoner. Den här regleringen är mycket viktig och en störning i funktionen hos dessa proteiner kan resultera i sjukdomar som till exempel cancer. Jag har studerat de proteindomäner från CREBBP och NCOA som möjliggör deras interaktion. Dessa domäner heter NCBD och CID, vilka båda hör till gruppen oordnade proteiner.

Ett av de stora målen i mina avhandlingsprojekt har varit att förstå hur interaktioner mellan proteiner kan förändras under evolutionens gång. För att förstå den här processen bättre har jag använt mig av olika versioner av NCBD och CID från ett tidsspann på flera hundra miljoner år, vilka har studerats med hjälp av olika biofysikaliska verktyg. Jag har till exempel tittat på hur dessa proteiners struktur och interaktionsmekanism har förändrats över detta långa tidsspann. Den samlade slutsatsen från dessa olika projekt är att interaktionen mellan NCBD och CID sannolikt har sett mycket lika ut överlag under många hundra miljoner år. Små förändringar i struktur och bindningsmekanism har uppstått vilka delvis kan förklara den skillnad i bindningsstyrka som har

observerats mellan de uråldriga proteinerna och nuvarande proteiner från människa och andra ryggradsdjur.

En senare del av mitt doktorandprojekt har gått ut på att studera hur oordnade proteinregioner som ligger utanför bindningsytan hos proteiner påverkar deras interaktion med andra proteiner. Jag har systematiskt tittat på hur omkringliggande regioner i CID påverkar interaktionen med NCBD. Resultaten visar att dessa regioner bidrar till en ökad bindningsstyrka för interaktionen. Effekten är inte beroende av en specifik proteinsekvens och verkar uppkomma som ett resultat av flera kortlivade ospecifika interaktioner. Dessa resultat demonstrerar hur de omkringliggande regionerna kan påverka slutsatserna när proteininteraktioner studeras och belyser vikten av att försöka inkludera större delar av proteinerna när experiment ska utföras.

Acknowledgements

My PhD supervisor **Per Jemth**, for letting me do my master thesis work in the group and later also admitted me to a PhD position. You have always been very encouraging and I am grateful for everything that I have learned in your lab.

Eva Andersson, for supervising me during my master thesis project, introducing me to the lab and teaching me about protein purification and other things. Lots of thanks for always making sure the lab is being run properly.

Emma Åberg and **Andreas Karlsson**, for making me feel welcome in the lab when I arrived as a master student and for making me feel like it was a really nice workplace. It was always a pleasure working with you and to share thoughts on everything in life.

Louise Laursen, we started our PhD in the lab at about the same time which make us some kind of scientific twins, I suppose. I admire your drive and strong will, and the way you were able to fit all the pieces of your busy life puzzle.

Filip Mihalic, I am so happy our PhD studies overlapped in the lab because a discussion with you is never boring. Your genuine curiosity about biology, physics and chemistry is very inspiring.

Weihua Ye, for always being a kind and friendly colleague.

Raviteja Inturi, for always being cheerful and fun to be around. For sharing parental joy and challenges, and for cooking delicious Indian food.

Amanda Lindberg, **Frieda Sorgenfrei**, **Zeinab Tehranizadeh** and **Mandy Erkelens**, it was a pleasure supervising all of you. You are all very smart and ambitious, and I hope your experience in the lab gave you lots of new knowledge and an appetite for more research.

Jan Schnatwinkel, for our collaboration on the flanking regions project. It was fun having you around and I learned a great deal from you as well.

Celestine Chi, for being a very generous and fun colleague. Special thanks for our collaborations and for introducing me to a tiny part of the NMR universe.

Magnus Kjaergaard, for our collaboration on the synchrotron-radiation CD paper.

Co-supervisor **Maria Selmer** and Examiner **Anna-Karin Olsson**, for the chats about research during our meetings.

The administration at IMBIM, that make IMBIM such a great environment for doing a PhD: **Alexis Fuentes** and **Tanel Punga**, for organizing the teaching schedule in the best way possible and always being reachable for questions and problem-solving. **Dorothe Spillmann**, for your devotedness to pedagogics and to making the wet-lab teaching sessions the best they can be. **Malin Strömbom**, **Malin Rask**, **Rehné Åkerblom** and **Veronica Hammar** for helping out with all kinds of matter. **Eva Gottfridsson**, for keeping the course labs organized and in exemplary shape from year to year.

Aida Paivandy, for all the time spent together at the enzyme kinetics lab. I am impressed by the way you took on the leadership role when starting to build up IPhA at IMBIM again. You set the barrier high from the beginning and created a stable foundation for the rest of us to build upon.

Viktor Ek, I am so impressed by your ambition and organization skills. You have contributed so much to the build-up of IPhA and for the social environment for PhD students at IMBIM. I hope you will make use of these skills in your future work role.

Isa Moreno, **Tifaine Hechard**, **Jennifer Jagdmann**, **Petra Geiser** and **Janina Krambrich**, for the time we spent together at different IPhA-related activities.

Johanna, I am so happy that you decided to come to our lab for a project and later also to stay in Uppsala for your PhD studies. Apart from being a really good friend and apart from all the good cakes that you bake, I am also very impressed by your PhD work. **Hanna**, **Mia** and **Malin**, it's always fun hanging out with you and hearing about your advancements in research, career and life.

"Umeåtjejerna" **Michelle**, **Sandra** och **Marika**, för kontakten vi har nu mestadels på distans, som ger mig så mycket glädje i vardagen, råd, peppning och stöttning. Det är så roligt att följa med på allt ni tar er an i livet.

Liselott och **Ole**, för att ni är så bra farföräldrar till Ralf och för att er dörr alltid är öppen för oss. **Fredriks mormor Siv**, för all god mat och avslappning hemma hos dig.

Mamma och **pappa**, för att jag är den jag är tack vare er. **Andreas**, för att du har ett så genuint intresse för att förstå hur saker fungerar. Jag är imponerad över hur många saker du har testat på och lärt dig bemästra. **Mormor** och **morfar**, för stödet jag får av er alltid.

Fredrik, du är den bästa människan som finns och en fantastisk pappa till Ralf.

Ralf, jag är så himla stolt över att vara din mamma. Du ger mig allt, och jag försöker ge dig allt jag kan tillbaka.

References

1. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).
2. Fersht, A. R. From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat. Rev. Mol. Cell Biol.* **9**, 650–654 (2008).
3. Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A. & Lansbury, P. T. NACP, A Protein Implicated in Alzheimer's Disease and Learning, Is Natively Unfolded †. *Biochemistry* **35**, 13709–13715 (1996).
4. Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. & Wright, P. E. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci.* **93**, 11504–11509 (1996).
5. Schweers, O., Schönbrunn-Hanebeck, E., Marx, A. & Mandelkow, E. Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. *J. Biol. Chem.* **269**, 24290–7 (1994).
6. Daughdrill, G. W., Chadsey, M. S., Karlinsey, J. E., Hughes, K. T. & Dahlquist, F. W. The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28. *Nat. Struct. Biol.* **4**, 285–91 (1997).
7. Oates, M. E. *et al.* D2P2: database of disordered protein predictions. *Nucleic Acids Res.* **41**, D508 (2013).
8. Demarest, S. J. *et al.* Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **415**, 549–553 (2002).
9. Moosa, M. M., Ferreon, J. C. & Ferreon, A. C. M. Ligand interactions and the protein order-disorder energetic continuum. *Semin. Cell Dev. Biol.* (2018). doi:10.1016/j.semcdb.2018.05.007
10. Fisher, C. K. & Stultz, C. M. Protein Structure along the Order–Disorder Continuum. *J. Am. Chem. Soc.* **133**, 10022–10025 (2011).
11. Yang, J., Gao, M., Xiong, J., Su, Z. & Huang, Y. Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding. *Protein Sci.* pro.3718 (2019). doi:10.1002/pro.3718
12. Dunker, A. K. *et al.* Intrinsically disordered protein. *J. Mol. Graph. Model.* **19**, 26–59 (2001).
13. Uversky, V. N. A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Sci.* **22**, 693–724 (2013).
14. Uversky, V. N., Gillespie, J. R. & Fink, A. L. Why are ?natively unfolded? proteins unstructured under physiologic conditions? *Proteins Struct. Funct. Genet.* **41**, 415–427 (2000).

15. Niklas, K. J., Dunker, A. K. & Yruela, I. The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *J. Exp. Bot.* **69**, 1437–1446 (2018).
16. Babu, M. M. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* **44**, 1185–1200 (2016).
17. Tompa, P., Davey, N. E., Gibson, T. J. & Babu, M. M. A million peptide motifs for the molecular biologist. *Mol. Cell* **55**, 161–9 (2014).
18. Van Roey, K., Gibson, T. J. & Davey, N. E. Motif switches: decision-making in cell regulation. *Curr. Opin. Struct. Biol.* **22**, 378–385 (2012).
19. Sundell, G. N. *et al.* Proteome-wide analysis of phospho-regulated PDZ domain interactions. *Mol. Syst. Biol.* **14**, e8129 (2018).
20. Babu, M. M. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* **44**, 1185–1200 (2016).
21. Schlessinger, A. *et al.* Protein disorder—a breakthrough invention of evolution? *Curr. Opin. Struct. Biol.* **21**, 412–418 (2011).
22. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Intrinsically Disordered Proteins in Human Diseases: Introducing the D ^{2} Concept. *Annu. Rev. Biophys.* **37**, 215–246 (2008).
23. Mollica, L. *et al.* Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Front. Mol. Biosci.* **3**, 52 (2016).
24. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8868–73 (2000).
25. Bagshaw, C. R. *Biomolecular Kinetics A Step-by-Step Guide*. (CRC Press, 2017).
26. Dyson, H. J. & Wright, P. E. Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300. *J. Biol. Chem.* **291**, 6714–22 (2016).
27. Kjaergaard, M., Andersen, L., Nielsen, L. D. & Teilum, K. A Folded Excited State of Ligand-Free Nuclear Coactivator Binding Domain (NCBD) Underlies Plasticity in Ligand Recognition. *Biochemistry* **52**, 1686–1693 (2013).
28. Kjaergaard, M., Poulsen, F. M. & Teilum, K. Is a Malleable Protein Necessarily Highly Dynamic? The Hydrophobic Core of the Nuclear Coactivator Binding Domain Is Well Ordered. *Biophys. J.* **102**, 1627–1635 (2012).
29. Demarest, S. J., Deechongkit, S., Dyson, H. J., Evans, R. M. & Wright, P. E. Packing, specificity, and mutability at the binding interface between the p160 coactivator and CREB-binding protein Specificity of binding between proteins using amphipathic helices is generally defined by the three-dimensional topology. *Gorina and Pavletich* **13**, 203–210 (1992).
30. Kjaergaard, M., Teilum, K. & Poulsen, F. M. Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. *Proc. Natl. Acad. Sci.* **107**, 12535–12540 (2010).

31. Haberz, P., Arai, M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Mapping the interactions of adenoviral E1A proteins with the p160 nuclear receptor coactivator binding domain of CBP. *Protein Sci.* **25**, 2256–2267 (2016).
32. Lee, C. W., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein. *Biochemistry* **49**, 9964–71 (2010).
33. Qin, B. Y. *et al.* Crystal Structure of IRF-3 in Complex with CBP. *Structure* **13**, 1269–1277 (2005).
34. Kjaergaard, M. *et al.* Temperature-dependent structural changes in intrinsically disordered proteins: Formation of α-helices or loss of polyproline II? *Protein Sci.* **19**, 1555–1564 (2010).
35. Hultqvist, G. *et al.* Emergence and evolution of an interaction between intrinsically disordered proteins. *Elife* **6**, (2017).
36. Iešmantavičius, V., Dogan, J., Jemth, P., Teilum, K. & Kjaergaard, M. Helical Propensity in an Intrinsically Disordered Protein Accelerates Ligand Binding. *Angew. Chemie Int. Ed.* **53**, 1548–1551 (2014).
37. Zosel, F., Mercadante, D., Nettels, D. & Schuler, B. A proline switch explains kinetic heterogeneity in a coupled folding and binding reaction. *Nat. Commun.* **9**, 3332 (2018).
38. Sturzenegger, F. *et al.* Transition path times of coupled folding and binding reveal the formation of an encounter complex. *Nat. Commun.* **9**, 4708 (2018).
39. Dogan, J., Schmidt, T., Mu, X., Engström, Å. & Jemth, P. Fast Association and Slow Transitions in the Interaction between Two Intrinsically Disordered Protein Domains. *J. Biol. Chem.* **287**, 34316–34324 (2012).
40. Matouschek, A., Kellis, J. T., Serrano, L. & Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122–126 (1989).
41. Johnson, C. M. Isothermal TitrationTitrations Calorimetry BT - Protein-Ligand Interactions: Methods and Applications. in (eds. Daviter, T., Johnson, C. M., McLaughlin, S. H. & Williams, M. A.) 135–159 (Springer US, 2021). doi:10.1007/978-1-0716-1197-5_5
42. Dogan, J., Mu, X., Engström, Å. & Jemth, P. The transition state structure for coupled binding and folding of disordered protein domains. *Sci. Rep.* **3**, 2076 (2013).
43. Jemth, P. *et al.* Structure and dynamics conspire in the evolution of affinity between intrinsically disordered proteins. *Sci. Adv.* **4**, (2018).
44. Ganguly, D., Zhang, W. & Chen, J. Synergistic folding of two intrinsically disordered proteins: searching for conformational selection. *Mol. Biosyst.* **8**, 198–209 (2012).
45. Karlsson, E. *et al.* Mapping the transition state for a binding reaction between ancient intrinsically disordered proteins. *J. Biol. Chem.* **295**, 17698–17712 (2020).
46. Kjaergaard, M., Teilum, K. & Poulsen, F. M. Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12535–40 (2010).

47. Eick, G. N., Bridgham, J. T., Anderson, D. P., Harms, M. J. & Thornton, J. W. Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty. *Mol. Biol. Evol.* **34**, 247–261 (2017).
48. Karlsson, E. *et al.* A structurally heterogeneous transition state underlies coupled binding and folding of disordered proteins. *J. Biol. Chem.* jbc.RA118.005854 (2018). doi:10.1074/jbc.RA118.005854
49. Wicky, B. I. M., Shammas, S. L. & Clarke, J. Affinity of IDPs to their targets is modulated by ion-specific changes in kinetics and residual structure. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9882–9887 (2017).
50. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583 (2021).

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1870

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-485228

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022