

## Minireview

# PreSMo Target-Binding Signatures in Intrinsically Disordered Proteins

Do-Hyoung Kim, and Kyou-Hoon Han\*

Genome Editing Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, 34141, Korea

\*Correspondence: khhan600@kribb.re.kr

<http://dx.doi.org/10.14348/molcells.2018.0192>[www.molcells.org](http://www.molcells.org)

Intrinsically disordered proteins (IDPs) are highly unorthodox proteins that do not form three-dimensional structures under physiological conditions. The discovery of IDPs has destroyed the classical structure-function paradigm in protein science, 3-D structure = function, because IDPs even without well-folded 3-D structures are still capable of performing important biological functions and furthermore are associated with fatal diseases such as cancers, neurodegenerative diseases and viral pandemics. Pre-structured motifs (PreSMos) refer to transient local secondary structural elements present in the target-unbound state of IDPs. During the last two decades PreSMos have been steadily acknowledged as the critical determinants for target binding in dozens of IDPs. To date, the PreSMo concept provides the most convincing structural rationale explaining the IDP-target binding behavior at an atomic resolution. Here we present a brief developmental history of PreSMos and describe their common characteristics. We also provide a list of newly discovered PreSMos along with their functional relevance.

**Keywords:** IDPs, IDR (Intrinsically Disordered Region), NMR, IUPs (Intrinsically Unfolded Proteins), PreSMos (Pre-Structured Motifs)

## INTRODUCTION

### Intrinsically Disordered Proteins

The central dogma in protein science, established over the

last half-century, states that “a well-folded 3-D structure is a prerequisite for protein function”. The 3-D structure in this statement refers to the one that is observed under near-physiological conditions, (i.e., ~ pH 7, ambient temperature, and aqueous buffer, etc.). Intrinsically unstructured/unfolded proteins (IUPs), now more commonly known as intrinsically disordered proteins (IDPs) (Dunker et al., 2013), are very peculiar proteins that do not form well-folded 3-D structures even under non-denaturing conditions. Naturally, IDPs are of great importance from a protein folding perspective. More intriguing are the observations that IDPs are functional or active without 3-D structures, for example, being involved in transcription (Lee et al., 2000; Sherr, 2004; Kim et al., 2017a; 2017b), translation (Fletcher and Wagner, 1998; Kim et al., 2015), cell cycle regulation (Pavletich, 1999), chaperoning (Hong et al., 2005), and membrane-binding (Atwal et al., 2007; Eliezer et al., 2001). The discovery of many, as much as half of the entire human proteome (Dunker et al., 2000), such highly unorthodox proteins has strongly suggested that the classical structure-function relationship of proteins needs to be reexamined. Clearly, the golden paradigm in structural biology, 3-D structure = protein function, is no longer valid. Several reviews dealing with general aspects of IDPs are available for further reading (Chavali et al., 2017; Dunker et al., 2013; Lee et al., 2012; Uversky and Dunker, 2010; Uversky, 2015).

Not only because of a basic scientific point of view are our interests in IDPs keen but also because of the fact that these proteins are involved in many fatal diseases. For example,

Received 2 May, 2018; revised 7 August, 2018; accepted 22 August, 2018; published online 10 October, 2018

eISSN: 0219-1032

© The Korean Society for Molecular and Cellular Biology. All rights reserved.

© This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>.

~80% of human cancers are associated with IDPs (Galea et al., 2008) such as eIF4E-binding proteins (4EBPs) (Fletcher and Wagner, 1998; Kim et al., 2015), Bcl-X<sub>L</sub> (Xu et al., 2009), human glucocorticoid receptors (Kim et al., 2017b), E7 (Lee et al., 2016), hypoxia inducible factors (Semenza, 2003; Kim et al., 2009a) and p53 all of which are so-called “hybrid-type” IDPs where intrinsically disordered regions (IDRs) coexist with globular domains (Lee et al., 2000; Wells et al., 2008). The causative agents of mad cow disease or Creutzfeldt-Jakob disease (CJD) in humans are prions that are also IDPs where a C-terminal globular domain coexists with a long intrinsically disordered region (IDR) at the N-terminus encompassing ~120 amino acid residues (James et al., 1997; Liu et al., 1999). Alpha-synuclein (Eliezer et al., 2001) and tau (Bibow et al., 2011; Künze et al., 2012), implicated in PD (Parkinson’s diseases) and AD (Alzheimer’s disease) respectively, are also IDPs. Furthermore, several viral strains including the well-known AIDS-causing HIV-1 produce IDPs (Chi et al., 2007; Feuerstein et al., 2012; Kim et al., 2009b; Lee et al., 2016; Liang et al., 2007; Reingewertz et al., 2009; To et al., 2016). Clearly, there is an immediate and strong need to acquire very thorough knowledge not only on the normal functionality of IDPs but also on their pathologic connection to above diseases since it has become apparent that the classical globular protein based approach is unlikely to provide us with sufficient information that can be used for developing effective weaponry against IDP-associated diseases.

### PreSMos: Pre-Structured Motifs, a Historical Perspective

The most obvious characteristic of IDPs is that they do not possess spatially-disposed active pockets, a fact that brings us to a simple but profound question of how then these long malleable stretches of amino acids (sometimes hundreds of amino acids) can bind their targets. Targets of IDPs are not just proteins, but can be nucleic acids (Thapar et al., 2004; To et al., 2016; Wells et al., 2008), lipids, metals, and small molecules (Follis et al., 2008; Metallo, 2010). Efforts were made recently to classify IDPs into several subfamilies (van der Lee et al., 2014). While intuitive, such a classification fails to provide detailed insights into how all these different subfamilies bind their targets. The well-cited expression “coupled folding and binding” (Dyson and Wright, 2002) is a useful term, but only as far as one tries to depict the rather easily-predictable topological change that IDPs need experience upon binding to their partners. This generic description therefore fails to provide any atomistic details associated with IDP-target binding that, if available, would be highly valuable for IDP-based drug design. As the axiom “*the devil is in the details*” dictates, the question one must answer is rather specific. It has been amply demonstrated that only certain segments or residues of IDPs/IDRs are involved in direct physical contact with target. Do we then have a clear answer on what specific features in these segments or residues make target binding possible? Why does mutating just a few (often just one) sparsely-disposed hydrophobic residues in acidic transactivation domains (TADs) drastically affect the transcriptional activity whereas mutating several of the abundant acidic residues has only a marginal effect on the activity? (Chang et al., 1995; Drysdale et

al., 1995) An early investigation attempted to address this question by employing wild type GAL4 and its scrambled mutant with no transcriptional activity (Giniger and Ptashne, 1987) and concluded that the mutant was inactive because its helix-forming propensity was compromised. This study triggered a huge controversy over whether target-free acidic TADs should form an amphipathic helix as the specificity determinant for activity (Van Hoy et al., 1993).

Direct and quantitative evidence that some sort of a secondary structural element, e.g., helix, is needed for transcriptional activity came from an NMR study on p53 TAD (Lee et al., 2000). The 73-residue long p53 TAD in its unbound form was found “unstructured” in a tertiary sense, yet contained a *transient* (~25% populated only) amphipathic helix whose residues formed a stable amphipathic helix when complexed with the N-terminal p53-binding domain (residues 3-109) of mdm2 (Kussie et al., 1996). This pioneering NMR study heralded the birth of the PreSMo concept. Subsequent NMR reports confirmed that pre-existing, pre-formed, or pre-ordered residual secondary structures, no matter what they may be called, do exist in unbound IDPs and are important for target binding (Lee et al., 2012). In the early days of IDP research, another line of thought prevailed advocating a notion of induced fit (IF), arguing that no pre-existing secondary structures were needed for target binding based upon the conclusion that IDPs are fully unstructured. A well-known example is the 4EBP1, a 118-residue translational inhibitor, which was reported to have “no regions of local order in the absence of eIF4E” (Fletcher et al., 1998). For the last two decades, this IDP has been known as the symbol of the completely unstructured (CU) nature of IDPs; however, a recent NMR study revealed that this IDP also contains a pre-structured helix which mediates its binding to eIF4E (Kim et al., 2015). Another well-known IDP is the kinase-inducible domain (KID) of CREB the NMR results on which supported the concept that IDPs must be in the CU state so that they must undergo “a coil → helix folding transition” via IF (Radhakrishnan et al., 1997). It is unclear how the authors of this particular report reached the conclusion that “the population of helix in free pKID is extremely small.” when their NMR data indicated presence of two transient helices (one ~60% and the other ~10% populated). Another group which worked on the same system concluded that two helix PreSMos were present (Table 1; Hua et al., 1998; Lee et al., 2012).

While the conceptual development on PreSMos has been somewhat delayed due to previous misconceptions that IDPs were completely unstructured, the presence of local residual secondary structures in isolated IDPs has been increasingly detected by many NMR investigations including a few critical NMR reports published at the turn of the century. The first key report found that p53 TAD has local structural elements (a helix and two turns) in the unbound state, as described above (Lee et al., 2000). The second report made by Ramelot et al. demonstrated that the cytoplasmic tail of the amyloid precursor protein forms a transient structure and such a pre-ordered structure is important for its binding to cytosolic factors (Ramelot et al., 2000). Sayers et al. also reported that structural preordering important for target binding was de-

tected in the N-terminal region of ribosomal protein S4 (Sayers et al., 2000). Zhao et al. reported local structural elements in the overall loosely folded Sml1 (Zhao et al., 2000). Zitzewitz et al. published an article in 2000 with a title of “Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system” (Zitzewitz et al., 2000). Another report by Bienkiewicz et al. described the functional consequences of pre-organized helical structure in the intrinsically disordered cell-cycle inhibitor p27 (Kip1) (Bienkiewicz et al., 2002). All these early NMR studies contributed to the foundation of the PreSMo concept, the idea that IDPs are not completely unstructured, but mostly unstructured (MU), and contain PreSMos. Following these NMR reports, bioinformatics studies proposed similar concepts such as PSE (Pre-formed Structural Element) (Fuxreiter et al., 2004), MoRF (Molecular Recognition Element) (Mohan et al., 2006; Oldfield et al., 2005), or primary contact sites a few years later. All these results, NMR experimental or predicted, point in unison to the idea that IDPs possess local secondary structural elements that are “hot spots” for target-binding.

In 2012 we published the first comprehensive review on PreSMos (Lee et al., 2012) because no explicit articles on the subject were available, despite the fact that PreSMos (whatever they may be called) have been recognized for more than a decade as very important (perhaps the most significant) features explaining IDP-target binding on a per-residue

basis. Several additional pieces of evidence have recently been published, demonstrating the functional significance of PreSMos (Kim et al., 2017b; lešmantavičius et al., 2014; Mohan et al., 2014; Salamanova et al., 2018). In the first review, we presented 27 IDPs/IDRs containing PreSMos which constitute ~56% of all IDPs characterized by then. Most critically, we introduced the term pre-structured motifs (PreSMos) in order to unambiguously point out the importance of the pre-structured nature of target-binding segments in free IDPs and to provide a convenient term that can replace various names “transient, nascent, residual, minimally-structured, non-negligible, pre-existing, pre-formed, or pre-ordered secondary structures”. These terms were used mainly by NMR structural biologists who did not hasten to generalize the concept with a particular name realizing that PreSMos had only been observed in a handful of IDPs until 2005. This review is a follow-up to our 2012 review. Because we have found 20 more PreSMos since our first review here we provide an updated list of PreSMos and a brief description on their functional significance; however, we acknowledge that the list may still be incomplete. In addition, we describe differences between the PreSMos that are detected experimentally and the terms derived from bioinformatics predictions. With this review we now have 47 IDPs/IDRs containing PreSMos, strongly suggesting that PreSMos are general *signatures* in most IDPs.

**Table 1.** A list of MU-type IDPs/IDRs containing PreSMos

Name	Number of residues	P/R <sup>b</sup>	Location of PreSMo residues <sup>c</sup>	Population <sup>d</sup> (%)	Role/Binding	References
FlgM	97	P	60-73 83-90 42-50	<b>50±10</b> <b>50±10</b> 20	$\sigma^{28}$	Daughdrill et al., 1997
KID	60	R	119-129 134-143	<b>&gt;50</b> <b>~10</b>	KIX	Radhakrishnan et al., 1998 Hua et al., 1998
GBD/CRIB in WASP W7	68 (201-268)	R	252-264	<b>~14</b>	Cdc42/Rac	Rudolph et al., 1998
HIV-1 Nef	56 (2-57)	R	14-22 : helix I 35-41 : helix II (H $\alpha$ only)	18 15		Geyer et al., 1999
Synaptobrevin-2	96	R	78-91	45	core complex forming	Hazzard et al., 1999
APPC	47 (649-695)	R	20-23 27-35 37-45 (H $\alpha$ only)	30 20 30	X11	Ramelot et al., 2000
p53 TAD	73	R	18-26 : helix 40-44 : turn I 48-53 : turn II	20 5 15	Mdm2 RPA, TFEII	Lee et al., 2000
RPS4	200	P	12-15 30-33: $\beta$ ?	8 23	rRNA, ribosomal proteins	Sayers et al., 2000
$\alpha$ -Synuclein	140	P	18-31	~10	amyloid-forming	Eliezer et al., 2001 Murrall et al., 2018
N-term. Tmod 1	92	R	24-35	NA	tropomyosin	Greenfield et al., 2005
VP16 TAD	79 (412-490)	R	443-447 469-483	25 15	hTAF <sub>31</sub> PC4	Jonker et al., 2005

(continued)

Name	Number of residues	P/R <sup>b</sup>	Location of PreSMo residues <sup>c</sup>	Population <sup>d</sup> (%)	Role/Binding	References
VP16 TAD	79 (412-490)	R	424-433/442-446, 465-467/472-479 (H $\alpha$ only)	60/40 10/20	hTAF <sub>II</sub> 31 PC4	Kim et al., 2009
Dynein interm. chain	40 (198-237)	R	223-228	NA	light chains	Benison et al., 2006 Benison et al., 2007
$\gamma$ -Synuclein	127	P	49-99	<b>~15</b>		Marsh et al., 2006
HMGA1	107	P	3-9 64-67	8	20 different proteins	Buchko et al., 2007
CFTR	185 (654-838)	R	<u><math>\alpha</math>-helix</u> 654-668, 759-764, 766-776, 801-817 <u><math>\beta</math>-strand</u> 744-753	<b>&gt;5</b> <b>&gt;5</b> <b>&gt;5</b>	interaction between R region and NT-binding domain 1	Baker et al., 2007
NS5A-D2 (HCV)	93 (250-342)	R	L48-V57 L86-E96 (H $\alpha$ only)	20 25	-	Liang et al., 2007
preS1 of HBV	119	R	32-36, 41-45 11-18, 22-25, 37-40, 46-50. (H $\alpha$ only)	~10 ~10 ~10	hepatocyte receptor-binding	Chi et al., 2007
$\beta$ -synuclein	134	P	NA	~20	-	Sung et al., 2007
Securin	202	P	150-159 : helix 113-127 ( $\beta$ ) 174-178	45 15 20	-	Csizmok et al., 2008
C-XPC <sup>e</sup>	126 (815-940)	R	818-843: helix 847-860: helix 891-901: helix 908-915: helix 923-930: helix	~30 ~30 NA NA NA	Centri2 TFIIH	Miron et al., 2008
MSP2	237	P	14-21 140-150 197-211	35 35 20	-	Zhang et al., 2008
DARPP-32	118	R	22-29 103-114	50 25	PP1	Dancheck et al., 2008
I-2	156 (9-164)	R	36-42 96-106 127-154 132-138	<b>30</b> <b>48</b> (70) <b>67</b> (90) <b>&gt;98</b>	PP1	Dancheck et al., 2008
ENSA	121	P	32-36 48-50 65-70	40 10 30	-	Boettcher et al., 2008 Boettcher et al., 2007
ODD/HIF-1 $\alpha$	74 (404-477)	R	438-440 467-477	~10	-	Kim et al., 2009
Sml1	104 (1-104)	P	4-14: helix 61-80: helix	~20 ~70	RNR binding Dimer forming	Zhao et al., 2000
Myb25	25 (291-315)	R	295-309 : helix	25~30	KIX	Zor et al., 2002
N tail Measles virus nucleoprotein	125 (401-525)	R	488-499 : helix	NA	phosphoprotein P	Bourhis et al., 2004
dSLBP	92 (17-108)	R	28-45 : helix 50-57 : helix 66-75 : helix 91-96 : helix	NA	mRNA stem-loop	Thapar et al., 2004
T $\beta$ -4	43 (1-43)	P	5-16 : helix	NA	Ca ATP G-actin	Domanski et al., 2004

(continued)

Name	Number of residues	P/R <sup>b</sup>	Location of PreSMo residues <sup>c</sup>	Population <sup>d</sup> (%)	Role/Binding	References
N tail	82	R	479-484	36	phosphoprotein P	<a href="#">Jensen et al., 2008</a>
Sendai Virus nucleoprotein	(443-524)		476-488	38		
			478-492	11		
Sic1	90	R	20-30	20	Cdc4	<a href="#">Mittag et al., 2008</a>
	(1-90)		63-68	20		
c-Myc	88	R	26-34 : helix	40	Bin-SH3 domain	<a href="#">Andresen et al., 2012</a>
	(1-88)		47-52 : helix	25	24-31(TRRAP binding)	
			20-23 : $\beta$ -turn			
ExsE	88	P	42-51: helix	NA	ExsC	<a href="#">Zheng et al., 2012</a>
	(1-88)		61-65: helix			
NS5A	415	R	401-412 : helix	NA	Bin1-SH3	<a href="#">Braeuning, 2013</a>
HCV	(33-447)		427-445 : helix			
NS5A	179	R	205-221 : helix I	38	Bin1-SH3	<a href="#">Feuerstein et al., 2012</a>
HCV	(191-369)		251-266 : helix II	38		<a href="#">Solyom et al., 2015</a>
			292-306 : helix III	51		
4EBP2	120	P	1-5	15~37	eIF4E	<a href="#">Lukhele et al., 2013</a>
	(1-120)		33-37			
			50-64			
			86-89			
			96-105			
E7	40	R	8-13 : helix	NA	E2	<a href="#">Noval et al., 2013</a>
HPV	(1-40)		17-29 : helix			
			33-38 : PPII			
4EBP1	70	R	56-63 : helix	20	eIF4E	<a href="#">Kim et al., 2015</a>
	(49-118)					
Myb32	32	R	290-310 : helix	~70	KIX	<a href="#">Arai et al., 2015</a>
	(284-315)					
E7	46	R	7-14 : helix	10	E2	<a href="#">Lee et al., 2016</a>
HPV	(1-46)		20-26 : helix	20		
CBP-ID4	207	R	1852-1875: helix	~60	-	<a href="#">Plai et al., 2016</a>
	(1851-2057)		1951-1978: helix			
HIV-1 Tat	121	P	27-32: helix	~20	Fab'	<a href="#">To et al., 2016</a>
	(1-121) <sup>a</sup>		41-59: helix	~30	P-TEFb	
			70-81: $\beta$ sheet	~25	TAR-cyclin T1	
			93-99: $\beta$ sheet	~25		
			105-112: $\beta$ sheet	~10		
SUSP4	100	R	263-291 : helix	~30	mdm2	<a href="#">Kim et al., 2017</a>
	(201-300)		265-270 : helix	~10		
			281-291 : helix			
hGRtau1c	64	R	185-202: helix	20~30	TAZ2	<a href="#">Kim et al., 2017</a>
	(181-244)		206-225: helix			
			232-244: helix			
Huntingtin Httex1 25Q	95	P	18-42: helix	NA	Cytotoxic	<a href="#">Newcombe et al., 2018</a>
	(1-95)				Membrane binding	
					Aggregation	

<sup>a</sup>The numbering includes a 20-residue N-terminal tag.

<sup>b</sup>An IDP (P) versus an IDR (R).

<sup>c</sup>Residue numbers are taken from the original report.

<sup>d</sup>Population of PreSMOs are read from the mid-point of the SSP scores that are calculated from chemical shifts in BMRB or literature. Shown in bold are the populations described in the original report. When the populations described in the original report without SSP scores differed significantly from the calculated SSP scores, the SSP scores are provided in parenthesis. NA = not available.

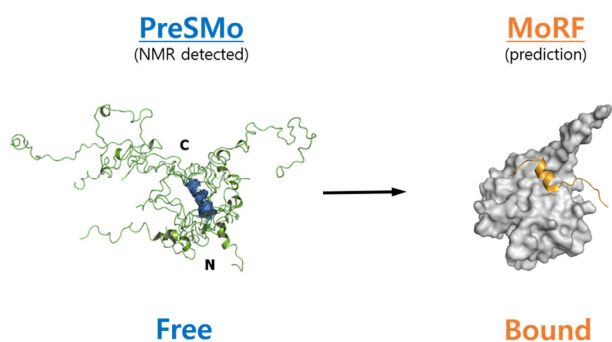
<sup>e</sup>Determined by SAXS.

## DISCUSSION

### Definition of a PreSMo

The definition of a PreSMo was given in our 2012 review (Lee et al., 2012); PreSMos are NMR-detected transient secondary structural elements within long (minimally 40 residues) and functionally-active IDRs of IDPs. We underline the fact that PreSMos are the *experimentally observable* entities in NMR analyses or other atomic-resolution experiments no matter how minimally it might be pre-populated; it is a measured quantity, not predicted notions. This contrasts with MoRF (Mohan et al., 2006), which is a theoretical concept derived from the target-bound conformations of short segments (peptides) of IDRs (Fig. 1). IDPs exist as an ensemble of many different conformers separated by small energy differences. A conformer with a PreSMo would be one in the ensemble that is populated to an NMR-detectable degree. The lowest population of a PreSMo-containing conformer observed to date is ~10% (Lee et al., 2012).

Table 1 is an updated list of PreSMos found in 47 IDPs/IDRs. The total number of IDPs studied in detail by NMR (with an exception of C-XPC studied by SAXS) is 70 even though the number of reports are more than 70 reports because some IDPs were investigated more than once. Notably, several IDPs (4EBP1, HIV-1 Tat, VP16 TAD, securin, and p21<sup>Waf1/Cipl/Sdil</sup>) that were originally reported as CU types with no PreSMo turned out to be MU types in later studies. For convenience, we added the 20 newly-identified PreSMos (starting from Myb25) at the end of Table 1, including a few PreSMos that were actually reported before 2012, but were not included in our 2102 review. Although the number of investigated IDPs is small compared to the possible number of IDPs/IDRs predicted by bioinformatics (thousands or more) it is sufficient



**Fig. 1. PreSMo vs. MoRF. A schematic diagram of the main differences between a PreSMo and a MoRF.** A PreSMo is observed mostly by NMR experiments in the target-free state of IDPs. Since free-state IDPs exist as an ensemble of many conformers separated by small energy differences, structural superposition among different conformers along the backbone atoms is not possible. Nevertheless, a structural superposition along a PreSMo is possible as shown in the left panel for the helix PreSMo of 4EBP1 (Kim et al., 2015). A PreSMo may become a MoRF upon target binding as illustrated for this helix PreSMo in 4EBP1 which becomes an  $\alpha$ -MoRF when bound to eIF4E.

to provide an overview on PreSMos. In 2012, the number of IDPs/IDRs with PreSMos was 27 (out of 48 studied) it is now 48 out of 70; the proportion of MU type IDPs/IDRs increased from 56% to 69%. The proportion is likely to increase if more IDPs/IDRs are characterized. One immediate feature noted in Table 1 is that in most cases we essentially study IDRs rather than IDPs (only 15 are IDPs), although we speak of IDPs. Note that all IDPs/IDRs in Table 1 are composed of more than 40 residues except for Myb25/Myb32. IDPs by definition consist of a minimal 40 residues and are distinct from the short flexible linkers and loops typically composed of fewer than 20 residues. The other feature shown in Table 1 is that most PreSMos are helices even though some are turns,  $\beta$ -strands and poly-proline type II helices. A high percentage of helices is also noted in MoRFs where  $\alpha$ -MoRFs are the majority (Mohan et al., 2006; Oldfield et al., 2005).

NMR is the main tool that enables quantitative definition of a PreSMo (Chi et al., 2007; Eliezer et al., 2001; Kim et al., 2009a; 2009b; 2015; 2017b; Lee et al., 2000; 2012; 2016; Liu et al., 1999; Xu et al., 2009). The beauty of NMR technique is that the presence of a PreSMo is reflected in several independent NMR parameters. In the early days, one needed to provide all of these NMR parameters (chemical shifts, inter-proton NOEs, J-couplings, T1 and T2 relaxation times, heteronuclear NOEs, temperature coefficients of backbone amide protons, etc.) to prove the existence of a PreSMo (Lee et al., 2000), whereas it usually is sufficient in recent years to just provide SSP (secondary structure propensity) scores (Marsh et al., 2006) as the concept of PreSMos has become more and more widely accepted. The SSP scores derived from CSIs (chemical shift indices) reveal an actual percentile value of a PreSMo population whereas CSIs can only indicate whether or not a PreSMo is present. A very important feature of a PreSMo is that it is never 100% populated. On the average, they are ~30% pre-populated, i.e., transient (Lee et al., 2012). This transient nature of PreSMos probably is the main cause that made several NMR investigators fail to detect them in the early days (Fletcher and Wagner, 1998; O'Hare and Williams, 1992; Radhakrishnan et al., 1997).

### PreSMo vs. MoRF

The most common bioinformatics term used interchangeably with PreSMos is MoRFs (Mohan et al., 2006). For example, the mdm2-binding helix PreSMo detected by NMR in free p53 TAD is reported as an  $\alpha$ -MoRF, a MoRF seen as an alpha helix in the target-bound state (Oldfield et al., 2005). Although there are a few more (out of more than a hundred) MoRFs that overlap with PreSMos fundamental differences exist between MoRFs and PreSMos. By definition MoRFs were identified in the x-ray structures of complexes between target proteins and short fragments of IDPs/IDRs that were predicted to be disordered by bioinformatics disorder prediction algorithms. The concept of the MoRF implicitly acknowledges the idea that the structured, bound-conformation is induced only upon target binding which is based on the early-day idea that IDPs have no pre-structured secondary structures. On the other hand, the definition of a PreSMo is not associated with the target-bound structure at all. In this regard, stating that a MoRF is found by NMR ex-

periments is inaccurate (Bourhis et al., 2004) since one cannot tell if a MoRF would exist within an isolated IDP. One has to obtain a complex structure between a target and a PreSMo/MoRF in order to conclude that the putative MoRF (which is actually a PreSMo) is indeed a MoRF. Thus, a helix PreSMo may become an  $\alpha$ -MoRF, but the opposite may not necessarily be true. With PreSMos we get the realistic percentage of the pre-structuredness whereas MoRFs do not provide such information. The term PreSMo was introduced as late as in 2012, but we underline that the PreSMos mentioned here refer to all the pre-existing or pre-formed residual secondary structures detected by NMR years before the term MoRF was introduced. It will be interesting to see how many of MoRFs may indeed coincide with PreSMos. One has to use a MoRF fragment, or preferably a longer IDR that encompasses such a MoRF fragment, to answer this question. An active pocket is a property of a globular protein that exists before binding to its target. In this regard, PreSMos qualify as the “active sites”, albeit not pockets, of IDPs since they are present before target binding. The same cannot be said for MoRFs. In Fig. 1, we show a conceptual scheme depicting what we have just described.

## Characteristics of PreSMos

### PreSMos are the “active sites” of IDPs

As is evident from Table 1 the PreSMos are the target-binding hot spots already present in free IDPs/IDRs; PreSMos are primed in a conformation similar to the target-bound conformation. Such pre-structuring is certainly advantageous for avoiding an entropic penalty that has to be paid when malleable IDPs/IDRs bind globular targets. Recent mutation studies demonstrated that the degree of pre-population of PreSMos is subtly controlled for efficient target binding (Borcherds et al., 2014; lešmantavičius et al., 2014; Kim et al., 2017b; Salamanova et al., 2018). In many globular proteins a single mutation in the active site completely nullifies protein function by disabling the binding of ligands. PreSMos are often found in tandem within sufficiently long transcription factor IDPs/IDRs separated by ~30 residues (Chi et al., 2005). One PreSMo may be a high-affinity binding site to a target whereas the other is a low-affinity site to the same target. A synergistic effect of multiple PreSMos for efficient target binding has been discussed previously (Lee et al., 2000).

### Shape complementarity in IDPs

Since it was believed that any secondary structure in IDPs should be induced only upon target binding many implicitly concluded that IDPs would totally lie outside of the classical structure-function paradigm, not obeying the rules established by structural biology such as shape complementarity. However, PreSMos reveal to us that IDPs abide by the shape complementarity extremely well via binding to targets (see Fig. 3 in Lee et al., 2012). In other words, when the secondary structural aspects for IDP-target binding are considered IDPs are not unorthodox at all. The genuine novelty of IDPs is the absence of 3-D structures only, not the absence of secondary structures. Structure (or PreSMos) does dictate func-

tion in the case of IDPs.

### Practical tips for NMR detection of PreSMos

The NMR spectral quality of hybrid-type IDPs is often not good enough for a full resonance assignment since a globular domain and an IDR will tumble around in different time scales. Consequently, a reductionist approach of using an IDR instead of a whole IDP is often necessary. One precaution when using such an approach is that one should use a sufficiently long region, not a short fragment since PreSMos may exist in the outside of the region covered by a short peptide (Botuyan et al., 1997; Uesugi et al., 1997). A longer IDR often contains a more populated PreSMo due to a tertiary effect that stabilizes the transient secondary structures, as was demonstrated in the case of p53 TAD and its short helical peptide (Botuyan et al., 1997; Lee et al., 2000). Another case demonstrating the significance of using a fragment of appropriate length is Myb 25/Myb32 (Table 1; Arai et al., 2015). The populations of a helix PreSMo in Myb25 and in Myb32 are ~30% and ~70%, respectively, demonstrating that having just 7 more residues in Myb32 drastically increases the PreSMo population by ~40%. Using bioinformatics disorder prediction programs may keep one from choosing an inappropriate IDR for NMR experiments. The inappropriate choice of an IDR for NMR investigation might be another reason why some NMR studies failed to detect PreSMos.

## CONCLUSION & PERSPECTIVE

Because IDPs are relatively a new field several new (sometimes rather vague) terms and expressions were introduced in order to describe novel concepts or phenomena associated with IDPs (van der Lee et al., 2014). Aside from bioinformatics terms (PSEs, MoRFs) other numerous expressions basically with the same meaning as PreSMos were proposed such as “only partly structured” (Zor et al., 2002), “small islands of secondary structures” (Laptenko and Prives, 2006), “weakly structured” (Chumakov, 2007), “limited structure” (Lavery and McEwan, 2008), “minimal ordering of short linear motifs” (Mittag et al., 2008), “residual secondary structural elements” (Kim et al., 2009b), “transient order” (Feuerstein et al., 2012), “transiently ordered regions”, “localized structurally ordered regions” (Zheng et al., 2012), and dynamic local structure (Lum et al., 2012) just to name a few.

Being flooded with so many terms that are intended to denote PreSMos is not unique for PreSMos. For example, it took more than a decade for the IDP research community to come up with a more or less consensus term for IDPs in 2013 (Dunker et al., 2013). Yet overly creative names not precisely in line with the classical concepts and terms in structural biology or protein science created a certain degree of confusion that led to a situation where the importance of IDPs was not duly appreciated for some time (Uversky and Dunker, 2010). Here, we present again an easy-to-use term of PreSMos to designate what has been described by several generic names realizing that the existence and functional significance of PreSMos will be appreciated more and more (now in ~70% of IDPs). Most importantly, the statement

that *IDPs would adopt structure only upon target binding* is misleading because it implies that IDPs are structureless down to the level of secondary structures. On the contrary, target binding only tightens (some structural induction) a PreSMo into a more stable conformation, but does not let a random-coil turn into a structure. In hindsight, the presence of PreSMos is in excellent agreement with the observations that a protein cannot exist in a *fully* random-coil state; denatured globular proteins are not random coils (Baldwin and Zimm, 2000; Bernadó et al., 2005; Neri et al., 1992).

Approximately 20 years have passed since IDPs emerged in protein science and structural biology communities. With more than ~5,000 papers on the subject no one would deny that IDPs have brought a critical paradigm shift to protein research, undoubtedly requiring that biochemistry textbooks be revised to include IDPs. There has been a tendency to put excessive emphasis on the disordered nature per se of IDPs with subsequent attempts trying to relate it to function due to an early-day misconception. For example, some reports on PreSMos were interpreted simply as evidence for disorder itself rather than as evidence for the existence of PreSMos (Cheng et al., 2006; Midic et al., 2009; Radivojac et al., 2007). It is important for the protein science community to learn a non-traditional view on proteins and their structures in two aspects. First, it is now well-known fact that long regions (40 residues and up) of proteins can be intrinsically disordered beyond the level of short disordered loops (Dunker et al., 2000). Proteins exist as dynamic conformational ensembles, not as snap-short entities that the PDB structures (both x-ray and NMR) have depicted for a long time. Second, in the absence of a well-defined 3D structure, the minimal residual secondary structures embedded into the flexible long IDR play key roles in target binding and govern the function of IDPs. Even in globular proteins, an important role of tertiary structure is to place the interacting (or active) secondary structures in a proper orientation relative to target proteins.

A discussion of PreSMos naturally brings us to the question of whether the mechanism of IDP-target binding follows IF (induced fit) or CS (conformational selection). In the case of KID-KIX binding IF (Sugase et al., 2007) was shown to be dominant whereas in the N-tail of viral nucleoproteins CS appeared prevalent (Jensen et al., 2008). In recent years, it is believed that these two mechanisms would work in concert: CS at the start of binding and IF at the final stage of binding (tightening). The existence of PreSMos itself is not an evidence for CS and one need to use a kinetics approach in order to determine if faster binding ( $k_{on}$  increased) can be achieved with more pre-structuring of the PreSMo segments. Future works employing PreSMo mutants should provide a more concrete answer on this aspect. No matter whether PreSMos are pre-structured or not, i.e., even if a PreSMo may become unstructured and re-structured for binding as one may envision in the IF model (To et al., 2016) it still does not change the fact that the fragment forming a PreSMo per se is important for target binding.

It is possible that PreSMos are also important for aggregation via oligomerization (Atwal et al., 2007; Eliezer et al., 2001). Both oligomerization and IDP-target binding are pro-

tein-protein interactions; the former is homogenous IDP-IDP self-binding while the latter is heterogeneous binding. Even though the PreSMo concept is broadly (~70%) applicable we do not expect that it should be applicable to all IDPs since there are IDPs/IDRs that are composed of simple dipeptide repeats (Lee et al., 2016). The PreSMo concept is also unlikely to be applicable to highly charged polyvalent IDPs which maintain unfolded topology even after target binding (Borgia et al., 2018). Due to strong attractive electrostatic interactions these IDPs have a very high affinity (pM) towards each other, unlike MU-type IDPs that bind their targets via PreSMos typically with  $\mu$ M affinities. However, it is noteworthy that even polyglutamine and polyproline were shown to form  $\alpha$ -helical and PPII helix type secondary structures, respectively (Mukrasch et al., 2009; Newcombe et al., 2018). Recent reports showed that IDP studies may lead to the development of new pharmaceuticals. For example, some PreSMo-antagonists against target proteins could serve as anti-cancer compounds (Kim et al., 2017a) and certain small molecule inhibitors can directly inhibit IDPs themselves (Follis et al., 2008; Metallo, 2010).

## ACKNOWLEDGMENTS

This work was supported by a Korea-Hungary and Pan EU collaborative project from National Research Council of Science and Technology (NST) (NTM2231712) to KH

## REFERENCES

- Andresen, C., Helander, S., Lemak, A., Farès, C., Csizmok, V., Carlsson, J., Penn, L.Z., Forman-Kay, J.D., Arrowsmith, C.H., Lundström, P., et al. (2012). Transient structure and dynamics in the disordered c-Myc transactivation domain affect Bin1 binding. *Nucleic Acids Res.* *40*, 6353-6366.
- Arai, M., Sugase, K., Dyson, H.J., and Wright, P.E. (2015). Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc. Natl. Acad. Sci. USA* *112*, 9614-9619.
- Atwal, R.S., Xia, J., Pinchev, D., Taylor, J., Epand, R.M., and Truant, R. (2007). Huntingtin has a membrane association signal that can modulate huntingtin aggregation, nuclear entry and toxicity. *Hum. Mol. Genet.* *16*, 2600-2615.
- Baker, J.M., Hudson, R.P., Kanelis, V., Choy, W.Y., Thibodeau, P.H., Thomas, P.J., and Forman-Kay, J.D. (2007). CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nat. Struct. Mol. Biol.* *14*, 738-745.
- Baldwin, R.L., and Zimm, B.H. (2000). Are denatured proteins ever random coils? *Proc Natl Acad Sci U S A.* *97*, 12391-12392.
- Benison, G., Nyarko, A., and Barbar, E. (2006). Heteronuclear NMR identifies a nascent helix in intrinsically disordered dynein intermediate chain: implications for folding and dimerization. *J. Mol. Biol.* *362*, 1082-1093.
- Benison, G., Berkholz, D.S., and Barbar, E. (2007). Protein assignments without peak lists using higher-order spectra. *J. Magn. Reson.* *189*, 173-181.
- Bernadó, P., Bertocini, C.W., Griesinger, C., Zweckstetter, M., and Blackledge, M. (2005). Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. *J. Am. Chem. Soc.* *127*, 17968-17969.
- Bibow, S., Mukrasch, M.D., Chinnathambi, S., Biernat, J., Griesinger,



- C., Mandelkow, E., and Zweckstetter, M. (2011). The Dynamic Structure of Filamentous Tau. *Angew. Chem. Int. Ed. Engl.* *50*, 11520-11524.
- Bienkiewicz, E.A., Adkins, J.N., and Lumb, K.J. (2002). Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1). *Biochemistry* *41*, 752-759.
- Boettcher, J.M., Hartman, K.L., Lador, D.T., Qi, Z., Woods, W.S., George, J.M., and Rienstra, C.M. (2008). Membrane-induced folding of the cAMP-regulated phosphoprotein endosulfine- $\alpha$ . *Biochemistry* *47*, 12357-12364.
- Boettcher, J.M., Hartman, K.L., Lador, D.T., Qi, Z., Woods, W.S., George, J.M., and Rienstra, C.M. (2007). (1)H, (13)C, and (15)N resonance assignment of the cAMP-regulated phosphoprotein endosulfine- $\alpha$  in free and micelle-bound states. *Biomol. NMR Assign.* *1*, 167-169.
- Borcherds, W., Theillet, F.X., Katzer, A., Finzel, A., Mishall, K.M., Powell, A.T., Wu, H., Manieri, W., Dieterich, C., Selenko, P., Loewer, A., and Daughdrill, G.W. (2014). Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nat. Chem. Biol.* *10*, 1000-1002.
- Borqia, A., Borqia, M.B., Buqqa, K., Kissling, V.M., Heidarsson, P.O., Fernandes, C.B., Sottini, A., Soranno, A., Buholzer, K.J., Nettels, D., Kragelund, B.B., Best, R.B., and Schuler, B. (2018). Extreme disorder in an ultrahigh-affinity protein complex. *Nature* *555*, 61-66.
- Bourhis, J.M., Johansson, K., Receveur-Bréchet, V., Oldfield, C.J., Dunker, K.A., Canard, B., and Longhi, S. (2004). The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon bind. to their physiological partner. *Virus Res.* *99*, 157-167.
- Botuyan, M.V., Momand, J., and Chen, Y. (1997). Solution conformation of an essential region of the p53 transactivation domain. *Fold Des.* *2*, 331-342.
- Braeuning, A. (2013). The connection of  $\beta$ -catenin and phenobarbital in murine hepatocarcinogenesis: a critical discussion of Awuah et al., *PLoS ONE* *7*, e39771. *Arch. Toxicol.* *87*, 401-402.
- Buchko, G.W., Ni, S., Lourette, N.M., Reeves, R., and Kennedy, M.A. (2007). NMR resonance assignments of the human high mobility group protein HMGA1. *J. Biomol. NMR.* *38*, 185.
- Chang, J., Kim, D.H., Lee, S.W., Choi, K.Y., and Sung, Y.C. (1995). Transactivation ability of p53 transcriptional activation domain is directly related to the binding affinity to TATA-binding protein. *J. Biol. Chem.* *270*, 25014-25019.
- Chavali, S., Gunnarsson, A., and Babu, M.M., (2017). Intrinsically disordered proteins adaptively reorganize cellular matter during stress. *Trends Biochem. Sci.* *42*, 410-412.
- Cheng, Y., LeGall, T., Oldfield, C.J., Dunker, A.K., and Uversky, V.N. (2006). Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry* *45*, 10448-10460.
- Chi, S.W., Lee, S.H., Kim, D.H., Ahn, M.J., Kim, J.S., Woo, J.Y., Torizawa, T., Kainosho, M., and Han, K. H. (2005). Structural details on mdm2-p53 interaction. *J. Biol. Chem.* *280*, 38795-38802.
- Chi, S.W., Kim, D.H., Lee, S.H., Chang, I., and Han, K.H. (2007). Pre-structured motifs in the natively unstructured preS1 surface antigen of hepatitis B virus. *Protein Sci.* *16*, 2108-2117.
- Chumakov, P.M. (2007). Versatile functions of p53 protein in multicellular organisms. *Biochemistry* *46*, 1399-1421.
- Csizmok, V., Felli, I.C., Tompa, P., Banci, L., and Bertini, I. (2008). Structural and dynamic characterization of intrinsically disordered human securin by NMR spectroscopy. *J. Am. Chem. Soc.* *130*, 16873-16879.
- Dancheck, B., Nairn, A.C., and Peti, W. (2008). Detailed structural characterization of unbound protein phosphatase 1 inhibitors. *Biochemistry* *47*, 12346-12356.
- Daughdrill, G.W., Chadsey, M.S., Karlinsey, J.E., Hughes, K.T., and Dahlquist, F.W. (1997). The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28. *Nat. Struct. Biol.* *4*, 285-291.
- Domanski, M., Hertzog, M., Coutant, J., Gutsche-Perelroizen, I., Bontems, F., Carlier, M.F., Guittet, E., and van Heijenoort, C. (2004). Coupling of folding and binding of thymosin beta4 upon interaction with monomeric actin monitored by nuclear magnetic resonance. *J. Biol. Chem.* *279*, 23637-23645.
- Drysdale, C.M., Dueñas, E., Jackson, B.M., Reusser, U., Braus, G.H., and Hinnebusch, A.G. (1995). The transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids. *Mol. Cell Biol.* *15*, 1220-1233.
- Dunker, A.K., Babu, M.M., Barbar, E., Blackledge, M., Bondosm, S.E., Dosztányi, Z., Dyson, H.J., Forman-Kay, J., Fuxreiter, M., Gsponer, J., Han, K.H., Jones, D.T., Longhi, S., Metallo, S.J., Nishikawa, K., Nussinov, R., Obradovic, Z., Pappu, R.V., Rost, B., Selenko, P., Subramaniam, V., Sussman, J.L., Tompa, P., and Uversky, V.N. (2013). What's in a name? Why these proteins are intrinsically disordered. Why these proteins are intrinsically disordered. *Intrinsically Disordered Proteins* *1*, e24157.
- Dunker, A.K., Obradovic, Z., Romero, P., Garner, E.C., and Brown, C.J. (2000). Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform.* *11*, 161-171.
- Dyson, H.J., and Wright, P.E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* *12*, 54-60.
- Eliezer, D., Kutluay, E., Bussell, R. Jr., and Browne, G., (2001). Conformational properties of  $\alpha$ -synuclein in its free and lipid-associated states. *J. Mol. Biol.* *307*, 1061-1073.
- Feuerstein, S., Solyom, Z., Aladag, A., Favier, A., Schwarten, M., Hoffmann, S., Willbold, D., and Brutscher, B. (2012). Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. *J. Mol. Biol.* *420*, 310-323.
- Fletcher, C.M., and Wagner, G. (1998). The interaction of eIF4E with 4E-BP1 is an induced fit to a completely disordered protein. *Protein Sci.* *7*, 1639-1642.
- Follis, A.V., Hammoudeh, D.I., Wang, H., Prochownik, E.V., and Metallo, S.J. (2008). Structural rationale for the coupled binding and unfolding of the c-myc oncoprotein by small molecules. *Chem. Biol.* *15*, 1149-1155.
- Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004). Prefolded structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* *338*, 1015-1026.
- Galea, C.A., Wang, Y., Sivakolundu, S.G., and Kriwacki, R.W., (2008). Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits. *Biochemistry* *47*, 7598-7609.
- Geyer, M., Munte, C.E., Schorr, J., Kellner, R., and Kalbitzer, H.R. (1999). Structure of the anchor-domain of myristoylated and non-myristoylated HIV-1 Nef protein. *J. Mol. Biol.* *289*, 123-138.
- Giniger, E., and Ptashne, M. (1987). Transcription in yeast activated by a putative amphipathic  $\alpha$  helix linked to a DNA binding unit. *Nature* *330*, 670-672.
- Greenfield, N.J., Kostyukova, A.S., and Hitchcock-DeGregori, S.E. (2005). Structure and tropomyosin binding properties of the N-terminal capping domain of tropomodulin 1. *Biophys. J.* *88*, 372-383.
- Hazzard, J., Südhof, T.C., and Rizo, J. (1999). NMR analysis of the structure of synaptobrevin and of its interaction with syntaxin. *J.*

Biomol. NMR. *14*, 203-207.

Hong, W., Jiao, W., Hu, J., Zhang, J., Liu, C., Fu, X., Shen, D., Xia, B., and Chang, Z. (2005). Periplasmic protein HdeA exhibits chaperone-like activity exclusively within stomach pH range by transforming into disordered conformation. *J. Biol. Chem.* *280*, 27029-27034.

Hua, Q.X., Jia, W.H., Bullock, B.P., Habener, J.F., and Weiss, M.A. (1998). Transcriptional activator-coactivator recognition: nascent folding of kinase-inducible transactivation domain predicts its structure on coactivator binding. *Biochemistry* *37*, 5858-5866.

lešmantavičius, V., Dogan, J., Jemth, P., Teilum, K., and Kjaergaard, M. (2014). Helical propensity in an intrinsically disordered protein accelerates ligand binding. *Angew. Chem. Int. Ed. Engl.* *53*, 1548-1551.

James, T.L., Liu, H., Ulyanov, N.B., Farr-Jones, S., Zhang, H., Donne, D.G., Kaneko, K., Groth, D., Mehlhorn, I., Prusiner, S.B., and Cohen, F.E. (1997). Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform. *Proc. Natl. Acad. Sci. USA* *94*, 10086-10091.

Jensen, M.R., Houben, K., Lescop, E., Blanchard, L., Ruigrok, R.W., and Blackledge, M. (2008). Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J. Am. Chem. Soc.* *130*, 8055-8061.

Jonker, H.R., Wechselberger, R.W., Boelens, R., Folkers, G.E., and Kaptein, R. (2005). Structural properties of the promiscuous VP16 activation domain. *Biochemistry* *25*, 827-839.

Kim, D.H., Lee, S.H., Chi, S.W., Nam, K.H., and Han, K.H. (2009a). Backbone resonance assignment of a proteolysis-resistant fragment in the oxygen-dependent degradation domain of the hypoxia inducible factor 1 $\alpha$ . *Mol. Cells* *27*, 493-496.

Kim, D.H., Lee, C., Lee, S.H., Kim, K.T., Han, J.J., Cha, E.J., Lim, J.E., Cho, Y.J., Hong, S.H., and Han, K.H. (2017a). The Mechanism of p53 Rescue by SUSP4. *Angew. Chem. Int. Ed. Engl.* *56*, 1278-1282.

Kim, D.H., Lee, S.H., Nam, K.H., Chi, S.W., Chang, I., and Han, K.H. (2009b). Multiple hTAF(II)31-binding motifs in the intrinsically unfolded transcriptional activation domain of VP16. *BMB Rep.* *42*, 411-417.

Kim, D.H., Lee, C., Cho, Y.J., Lee, S.H., Cha, E.J., Lim, J.E., Sabo, T.M., Griesinger, C., Lee, D., and Han, K.H. (2015). A pre-structured helix in the intrinsically disordered 4EBP1. *Mol. Biosyst.* *11*, 366-369.

Kim, D.H., Wright, A., and Han, K.H. (2017b). An NMR study on the intrinsically disordered core transactivation domain of human glucocorticoid receptor. *BMB Rep.* *10*, 522-527.

Künze, G., Barré, P., Scheidt, H.A., Thomas, L., Eliezer, D., and Huster, D. (2012). Binding of the three-repeat domain of tau to phospholipid membranes induces an aggregated-like state of the protein. *Biochim. Biophys. Acta.* *1818*, 2302-2313.

Kussie, P.H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A.J., and Pavletich, N.P. (1996). Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* *274*, 948-953.

Laptenki, O., and Prives, C. (2006). Transcriptional regulation by p53: one protein, many possibilities. *Cell Death Differ.* *13*, 951-961.

Lavery, D.N., and McEwan, I.J. (2008) Structural characterization of the native NH2-terminal transactivation domain of the human androgen receptor: a collapsed disordered conformation underlies structural plasticity and protein-induced folding. *Biochemistry* *47*, 3360-3369.

Lee, C., Kim, D.H., Lee, S.H., Su, J., and Han, K.H. (2016). Structural investigation on the intrinsically disordered N-terminal region of HPV16 E7 protein. *BMB Rep.* *49*, 431-436.

Lee, H., Mok, K.H., Muhandiram, R., Park, K.H., Suk, J.E., Kim, D.H., Chang, J., Sung, Y.C., Choi, K.Y., and Han, K.H. (2000). Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *J. Biol. Chem.* *275*, 29426-29432.

Lee, K.H., Zhang, P., Kim, H.J., Mitrea, D.M., Sarkar, M., Freibaum, B.D., Cika, J., Coughlin, M., Messing, J., Molliex, A., Maxwell, B.A., Kim, N.C., Temirov, J., Moore, J., Kolaitis, R.M., Shaw, T.I., Bai, B., Peng, J., Kriwacki, R.W., and Taylor, J.P. (2016). C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell* *167*, 774-788.

Lee, S.H., Kim, D.H., Han, J.J., Cha, E.J., Lim, J.E., Cho, Y.J., Lee, C., and Han, K.H. (2012). Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins. *Curr. Protein Pept. Sci.* *13*, 34-54.

Liu, H., Farr-Jones, S., Ulyanov, N.B., Llinas, M., Marqusee, S., Groth, D., Cohen, F.E., Prusiner, S.B., and James, T.L. (1999). Solution Structure of Syrian Hamster Prion Protein rPrP(90-231). *Biochemistry* *38*, 5362-5377.

Liang, Y., Ye, H., Kang, C.B., and Yoon, H.S. (2007). Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. *Biochemistry* *46*, 11550-11558.

Lukhele, S., Bah, A., Lin, H., Sonenberg, N., and Forman-Kay, J.D. (2013). Interaction of the eukaryotic initiation factor 4E with 4E-BP2 at a dynamic bipartite interface. *Structure* *21*, 2186-2196.

Lum, J.K., Neuweiler, H., and Fersht, A.R. (2012). Long-range modulation of chain motions within the intrinsically disordered transactivation of tumor suppressor p53. *J. Am. Chem. Soc.* *124*, 1617-1622.

Marsh, J.A., Singh, V.K., Jia, Z., and Forman-Kay, J.D. (2006). Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: Implications for fibrillation. *Protein Sci.* *15*, 2795-2804.

Metallo, S.J. (2010). Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* *14*, 481-488.

Midic, U., Oldfield, C.J., Dunker, A.K., Obradovic, Z., and Uversky, V.N. (2009). Protein disorder in the human diseasome: unfoldomics of human genetic diseases. *BMC Genomics.* *10*, S12.

Miron, S., Duchambon, P., Blouquit, Y., Durand, D., and Craescu, C.T. (2008). The carboxy-terminal domain of xeroderma pigmentosum complementation group C protein, involved in TFIIH and centrin binding, is highly disordered. *Biochemistry* *47*, 1403-1413.

Mittag, T., Orlicky, S., Choy, W.Y., Tang, X., Lin, H., Sicheri, F., Kay, L.E., Tyers, M., and Forman-Kay, J.D. (2008). Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc. Natl. Acad. Sci. USA* *105*, 17772-17777.

Mohan, A., Oldfield, C.J., Radivojac, P., Vacic, V., Cortese, M.S., Dunker, A.K., and Uversky, V.N. (2006). Analysis of Molecular Recognition Features (MoRFs). *J. Mol. Biol.* *362*, 1043-1059.

Mukrasch, M.D., Bibow, S., Korukottu, J., Jeganathan, S., Biernat, J., Griesinger, C., Mandelkow, E., and Zweckstetter, M. (2009). Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol.* *7*, e34.

Murali, M.G., Schiavina, M., Sainati, V., Bermel, W., Pierattelli, R., and Felli, I.C. (2018). <sup>13</sup>C APSY-NMR for sequential assignment of intrinsically disordered proteins. *J. Biomol. NMR.* *70*, 167-175.

Neri, D., Billeter, M., Wider, G., and Wüthrich, K. (1992). NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* *257*, 1559-1563.

Newcombe, E.A., Ruff, K.M., Sethi, A., Ormsby, A.R., Ramdzan, Y.M., Fox, A., Purcell, A.W., Gooley, P.R., Pappu, R.V., and Hatters, D.M. (2018). Tadpole-like conformations of huntingtin exon 1 are

- characterized by conformational heterogeneity that persists regardless of polyglutamine length. *J. Mol. Biol.* *430*, 1442-1458.
- Noval, M.G., Gallo, M., Perrone, S., Salvay, A.G., Chemes, L.B., and de Prat-Gay, G. (2013). Conformational dissection of a viral intrinsically disordered domain involved in cellular transformation. *PLoS One* *8*, e72760.
- O'Hare, P., and Williams, G. (1992). Structural studies of the acidic transactivation domain of the Vmw65 protein of herpes simplex virus using <sup>1</sup>H NMR. *Biochemistry* *31*, 4150-4156.
- Oldfield, C.J., Cheng, Y., Cortese, M.S., Romero, P., Uversky, V.N., and Dunker, A.K. (2005). Coupled folding and binding with alpha-helix forming molecular recognition elements. *Biochemistry* *44*, 12454-12470.
- Pavletich, N.P. (1999). Mechanisms of cyclin-dependent kinase regulation: structures of cdk's, their cyclin activators, and cip and INK4 inhibitors. *J. Mol. Biol.* *287*, 821-828.
- Piai, A., Calçada, E.O., Tarenzi, T., Grande, A.D., Varadi, M., Tompa, P., Felli, I.C., and Pierattelli, R. (2016). Just a Flexible Linker? The structural and dynamic properties of CBP-ID4 revealed by NMR spectroscopy. *Biophys. J.* *110*, 372-381.
- Radivojac, P., Iakoucheva, L.M., Oldfield, C.J., Obradovic, A., Uversky, V.N., and Dunker, A.K. (2007). Intrinsic disorder and functional proteomics. *Biophys. J.* *92*, 1439-1456.
- Radhakrishnan, I., Pérez-Alvarado, G.C., Parker, D., Dyson, H.J., Montminy, M.R., and Wright, P.E. (1997). Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* *91*, 741-752.
- Radhakrishnan, I., Pérez-Alvarado, G.C., Dyson, H.J., and Wright, P.E. (1998). Conformational preferences in the Ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS Lett.* *430*, 317-322.
- Ramelot, T.A., Gentile, L.N., and Nicholson, L.K. (2000). Transient structure of the amyloid precursor protein cytoplasmic tail indicates preordering of structure for binding to cytosolic factors. *Biochemistry* *39*, 2714-2725.
- Reingewertz, T.H., Benyamini, H., Lebendiker, M., Shalev, D.E., and Friedler, A. (2009). The C-terminal domain of the HIV-1 Vif protein is natively unfolded in its unbound state. *Protein Eng. Des. Sel.* *22*, 281-287.
- Rudolph, M.G., Bayer, P., Abo, A., Kuhlmann, J., Vetter, I.R., and Wittinghofer, A. (1998). The Cdc42/Rac interactive binding region motif of the Wiskott Aldrich syndrome protein (WASP) is necessary but not sufficient for tight binding to Cdc42 and structure formation. *J. Biol. Chem.* *273*, 18067-18076.
- Salamanova, E., Costeira-Paulo, J., Han, K.H., Kim, D.H., Nilsson, L., and Wright, A.P.H. (2018). A subset of functional adaptation mutations alter propensity for  $\alpha$ -helical conformation in the intrinsically disordered glucocorticoid receptor tau1core activation domain. *Biochim. Biophys. Acta.* *1862*, 1452-1461.
- Sayers, E.W., Gerstner, R.B., Draper, D.E., and Torchia, D.A. (2000). Structural preordering in the N-terminal region of ribosomal protein S4 revealed by heteronuclear NMR spectroscopy. *Biochemistry* *39*, 13602-13613.
- Semenza, G.L. (2003). Targeting HIF-1 for cancer therapy. *Nat. Rev. Cancer.* *3*, 721-732.
- Sherr, C.J. (2004). Principles of tumor suppression. *Cell* *116*, 235-246.
- Sólyom, Z., Ma, P., Schwarten, M., Bosco, M., Polidori, A., Durand, G., Willbold, D., and Brutscher, B. (2015). The Disordered Region of the HCV Protein NS5A: Conformational Dynamics, SH3 Binding, and Phosphorylation. *Biophys. J.* *109*, 1483-1496.
- Sugase, K., Dyson, H.J., and Wright, P.E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* *447*, 1021-1025.
- Sung, Y.H., and Eliezer, D. (2007). Residual structure, backbone dynamics, and interactions within the synuclein family. *J. Mol. Biol.* *372*, 689-707.
- Thapar, R., Mueller, G.A., and Marzluff, W.F. (2004). The N-terminal domain of the Drosophila histone mRNA binding protein, SLBP, is intrinsically disordered with nascent helical structure. *Biochemistry* *43*, 9390-9400.
- To, V., Dzananovic, E., McKenna, S.A., and O'Neil, J. (2016). The Dynamic Landscape of the Full-Length HIV-1 Transactivator of Transcription. *Biochemistry* *55*, 1314-1325.
- Uesugi, M., Nyanguile, O., Lu, H., Levine, A.J., and Verdine, G.L. (1997). Induced alpha helix in the VP16 activation domain upon binding to a human TAF. *Science* *277*, 1310-1313.
- Uversky, V.N. (2015). Functional roles of transiently and intrinsically disordered regions within proteins. *FEBS J.* *282*, 1182-1189.
- Uversky, V.N., and Dunker, A.K. (2010). Understanding protein non-folding. *Biochim. Biophys. Acta.* *1804*, 1231-1264.
- van der Lee, R., Buljan, M., Lanq, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Krivacki, R.W., Oldfield, C.J., Pappu, R.V., Tompa, P., Uversky, V.N., Wright, P.E., and Babu, M.M. (2014). Classification of intrinsically disordered regions and proteins. *Chem. Rev.* *114*, 6589-6631.
- Van Hoy, M., Leuther, K.K., Kodadek, T., and Johnston, S.A. (1993). The acidic activation domains of the GCN4 and GAL4 proteins are not alpha helical but form beta sheets. *Cell* *72*, 587-594.
- Wells, M., Tidow, H., Rutherford, T.J., Markwick, P., Jensen, M.R., Mylonas, E., Svergun, D.I., Blackledge, M., and Fersht, A.R. (2008). Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. USA* *105*, 5762-5767.
- Xu, H., Ye, H., Osman, N.E., Sadler, K., Won, E.Y., Chi, S.W., and Yoon, H.S. (2009). The MDM2-binding region in the transactivation domain of p53 also acts as a Bcl-X(L)-binding motif. *Biochemistry* *48*, 12159-12168.
- Zhang, X., Perugini, M.A., Yao, S., Adda, C.G., Murphy, V.J., Low, A., Anders, R.F., and Norton, R. S. (2008). Solution conformation, backbone dynamics and lipid interactions of the intrinsically unstructured malaria surface protein MSP2. *J. Mol. Biol.* *379*, 105-121.
- Zhao, X., Georgieva, B., Chabes, A., Domkin, V., Ippel, J.H., Schleucher, J., Wijmenga, S., Thelander, L., and Rothstein, R. (2000). Mutational and structural analyses of the ribonucleotide reductase inhibitor Sml1 define its Rnr1 interaction domain whose inactivation allows suppression of mec1 and rad53 lethality. *Mol. Cell Biol.* *23*, 9076-9083.
- Zheng, Z., Ma, D., Yahr, T.L., and Chen, L. (2012). The transiently ordered regions in intrinsically disordered ExsE are correlated with structural elements involved in chaperone vinding. *Biochem. Biophys. Res. Commun.* *417*, 129-134.
- Zitzewitz, J.A., Ibarra-Molero, B., Fishel, D.R., Terry, K.L., and Matthews, C.R. (2000). Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J. Mol. Biol.* *296*, 1105-1116.
- Zor, T., Mayr, B.M., Dyson, H.J., Montminy, M.R., and Wright, P.E. (2002). Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. *J. Biol. Chem.* *277*, 42241-42248.