Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

A pre-structured helix in the intrinsically disordered 4EBP1

Do-Hyoung Kim[a‡], Chewook Lee[a‡], Ye-Jin Cho[a,b‡], Si-Hyung Lee[a], Eun-Ji Cha[a], Ji-Eun Lim[a], T. Michael Sabo[c], Christian Griesinger[c], Donghan Lee[c], and Kyou-Hoon Han[a,b]*

Received (in XXX, XXX) XthXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

The eIF4E-binding protein 1 (4EBP1) has long been known to be completely unstructured without any secondary structures, which contributed significantly to the proposal of the induced fit mechanism for target binding of intrinsically disordered proteins. We show here that 4EBP1 is not completely unstructured, but contains a pre-structured helix.

Intrinsically disordered proteins (IDPs) are associated with broad biological functions as well as with critical diseases including prion ("mad cow") diseases, cancers, viral infection and neurodegenerative diseases[1-5]. As the eventual function of most (~80%) IDPs is to convey biological signals by binding to various types of target molecules, such as proteins, nucleic acids, metals, or lipids[1,6,7], delineating their target-binding mechanism is important to clearly understand IDP function. Recent studies illustrate that accurate structural knowledge on IDPs may have immediate consequences even for drug development[4,5]. A disorder to order transition and coupled folding and binding are common terms describing IDP-target binding[2,8-10]. These terms, however, mostly refer to a global topological change occurring in IDPs upon target binding. At an atomistic level an induced fit (IF) mechanism involving a coil → helix transition was proposed at the dawn of the IDP research arguing that any pre-structuring of the target-binding segment is unnecessary for binding[8-11]. However, a coil → helix structural transition is not likely to occur if a target-binding segment in a free IDP is already pre-structured in a conformation that presages its target-bound conformation[6]**.** In such a case conformational selection of the pre-structured segment by a target may be an efficient and more thermodynamically favorable event. Thus, a fundamental question concerns whether IDPs in their free state are totally unstructured down to the level of secondary structures[12-14] noting that even fully denatured globular proteins cannot be described by a complete random coil model[15].

A recent analysis on ~50 IDPs and IDRs (intrinsically disordered regions) whose conformational details were characterized by NMR techniques revealed that ~70% of them are in a mostly unstructured (MU) state rather than being in a completed unstructured (CU) state[6]. The MU-type IDPs contain the so-called **pre-s**tructured **mo**tif**s** (PreSMos), originally coined as a **lo**cal **st**ructural (***lost***) elements[3], almost all of which serve as the specific determinants for target binding. After the introduction of the PreSMo concept several CU type IDPs originally proposed to undergo the coil → helix IF transition were carefully re-analyzed by NMR and turned out to be MU-types, seriously weakening the basis knowledge supporting the coil → helix IF mechanism[6]. These results pointed out a need to rekindle the early idea on the potential contribution of conformational selection of a PreSMo by a target protein to IDP-target binding[3]. Nonetheless, the IF mechanism has been mostly considered in the IDP field. Whilst presence of a PreSMo *per se* certainly is not a sufficient condition for conformational selection it seems clear that the subtly controlled level - neither too little nor too much- of secondary structure pre-population of the target-binding segments in free IDPs is important for target binding[16]. For example, a recent mutation study on an IDR of thyroid hormone and retinoid receptors (ACTR) showed that the helical fraction of a helical PreSMo in the unbound ACTR correlated with its binding affinity to the nuclear coactivator binding domain (NCBD) of the CREB binding protein[17]. Early reports also pointed out the pre-structuring of the target-binding segments[6,16] .

The human phosphoprotein 4EBP1 is the very first IDP *explicitly* described to be completely or "wholly" disordered[8,9], which contributed critically to the formation of a coil → helix IF concept. Interestingly, this paradigmatic IDP was not re-analyzed in the context of the PreSMo concept. The 4EBP1 contains an eIF4E-binding segment composed of residues 55-63[2]. Given that PreSMos are target-binding motifs[6] we postulated that the residues 55-63 in 4EBP1 form a PreSMo. The early NMR data on 4EBP1 did not contain a complete resonance assignment due to resonance overlap[8]. In order to overcome this overlap problem we used a shorter construct of 4EBP1 (residues 49-118; named BP49 hereafter) encompassing the eIF4E-binding region.

Chemical shifts (**Fig. 1a** and **1b**) are the first NMR parameters to be used to determine if an IDP contains a PreSMo[6]. The SSP score of ~0.2 in BP49 (**Fig. 1c**) indicates that the eIF4E-binding residues 56-63 adopt ~20% of a helix in a free state. A similar degree of pre-population is noted for many PreSMos[6]. Existence of this helix PreSMo is also supported by the backbone dynamics (**Fig. 1d**); positive values (0.3~0.5) of ^{1}H-^{15}N heteronuclear NOEs are observed for these residues although they are not as large as those (0.8~1.0) obtained for a stable helix (**Fig 1e**). The ^{15}N relaxation times, particularly T_2, for the PreSMo-forming residues clearly deviate from the rest of the molecule as indicated by the J(0) values ranging between 2 and 2.7 rad/nsec indicating somewhat restricted motion (**Fig. S5, ESI†**). Contiguously observed small temperature coefficients (< 5 ppb/K) of the backbone amide NHs (residues 56-63) (**Fig. 1d**) also suggest formation of a helix.

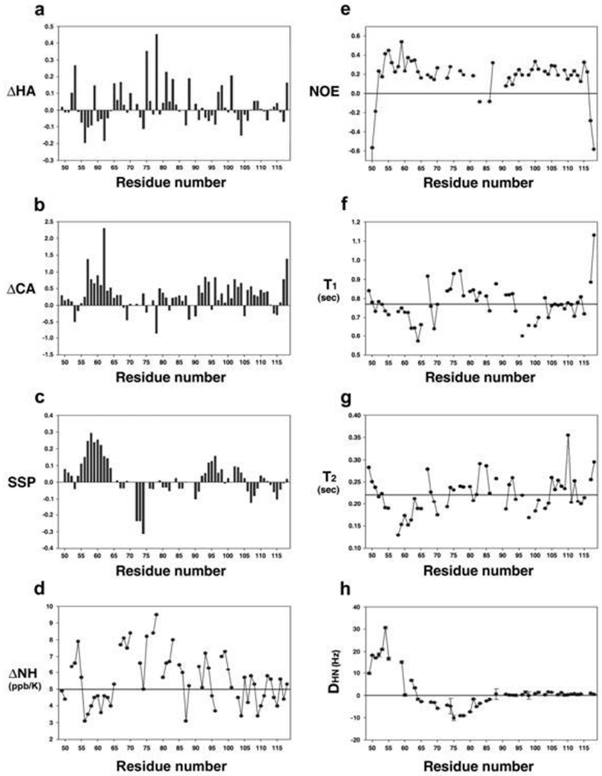

Figure 1. Left panel: Deviation of ^{1}Hα (**a**) and ^{13}Cα (**b**) chemical shifts from random coil values. The SSP (secondary structure propensity) scores (**c**) and temperature coefficients of the backbone amide hydrogens (**d**). Right panel: ^{1}H-^{15}N heteronuclear NOEs (**e**) and backbone ^{15}N relaxation times, T_1 (**f**) and T_2 (**g**), and NH residual dipolar coupling constants (**h**) of BP49. The horizontal lines in (**f**) and (**g**) indicate an average.

We further characterized BP49 using the Flexible-Meccano (FM) approach[18] to determine the content of PreSMo. Experimental residual dipolar couplings (RDCs) measured under negatively charged Pf-1 phages (**Fig. S1, ESI†)** were used to generate an ensemble structure for IDPs. **Figure 1h** shows that the N-terminus of BP49 displays RDC values deviating from a completely disordered segment. The FM approach predicts that ~15% of BP49 is engaged in helix formation similar to that obtained from SSP. FM ensembles yield two helices between residues 57-62 (6.6 ± 0.2%) and 51-60 (8.3 ± 0.1%) (**Fig.S2 and Table S1, ESI†**) and their presence is also supported by the experimental observation of interproton NOEs for these helices (**Fig.S3 and S4, ESI†**). In addition when a conformational ensemble of BP49 is calculated by replica exchange molecular dynamics (REMD) is the residues 56-63 are shown to form a helix. **In Figure 2a** and **2b** we present 10 REMD ensemble structures of BP49 in the eIF4E-free state and the x-ray structure of an eIF4E-bound 4EBP1 peptide[2]. **Figure 2c** illustrates how remarkably the pre-structured helix presages the eIF4E-bound helix. The REMD ensemble reveals two H-bonds formed at the N-terminus of the pre-structured helix between the side chain carboxylate group of 55D and the backbone NHs of 56R and 57K (see **Fig. 3d**) in agreement with the small temperature coefficients. (**Fig. 2d**).

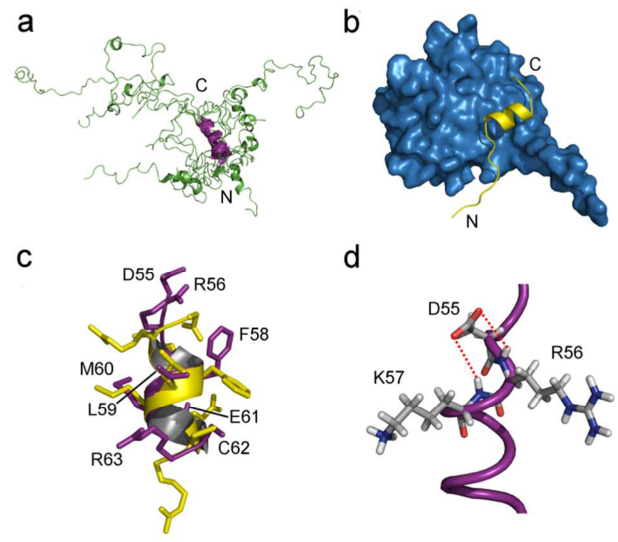

Figure 2. REMD ensemble of BP49 superimposed over a pre-structured helix (purple) (**a**), the x-ray structure of an eIF4E-bound 4EBP1 peptide (residues 51-67) (yellow)[2] (**b**), and superposition of the pre-structured helix (purple) and the eIF4E-bound helix (yellow) (**c**). Two N-terminal hydrogen bonds involving the side chain carboxylate group of 55D and the backbone amide protons of 56R and 57K (**d**).

The observations that IDPs or IDRs containing long (> 40 residues) disordered segments could carry out inherent functions, e.g, transcription and translation, without using 3-D structures were novel enough to generate a serious query on their target recognition process[3,12-15]. The rationale that IDPs, being fundamentally different from globular proteins, may well have their own unique mechanism of target binding that defies the conformational complementarity rule globular proteins obey seemed acceptable to a certain degree. However, such an explanation is not sufficient in answering an unavoidable question, "How a protein, no matter how novel they may be, could recognize its targets in such a non-specific way (i.e. without a 3-D structure) without relying on conformational traits at all?". Note that this question applies to ~80% of IDPs[7]. Within this context the discovery of a PreSMo as an "active site" in the intrinsically disordered transactivation domain (TAD) of p53 was rather revealing since it demonstrated that local secondary structural elements in free IDPs could be the answer to the above question[3,6]. In fact, we now realize that IDPs are not total outliers *completely* defying the classical structure-function paradigm in the protein kingdom because IDPs use PreSMos to abide by the shape complementarity rule[6].

The PreSMo concept was poorly recognized in the early days when a few reports described that IDPs were in a CU state[8-11]. One of these studies involved a short fragment (residues 469-482) in the VP16 TAD putatively undergoing a coil-to-helix IF; the helix formed in the $TAF_{II}31$-bound state of VP16 TAD was not observed in the unbound state[11]. However, three independent NMR studies using a longer segment of VP16 TAD showed later that the putative segment formed a helix PreSMo[6]. A transient secondary structure in a short peptide can be easily missed if studied in aqueous solution in isolation unless it has an extremely strong inherent propensity to form a secondary structure[19]. The reductionistic approach of using a short VP16 TAD peptide seems to have led to an erroneous conclusion that the putative segment of VP16 TAD underwent a coil-to-helix IF. Another misleading report dealt with a sufficiently long (~60 residues) KID fragment of CREB. Somehow this IDR was described to contain "extremely small" fraction of secondary structures, which inevitably supported a "coil-to-helix" IF[9] when in fact as shown in a later study[16,20] that the free KID was populated with two helix PreSMos, one pre-structured at ~50% and the other T at ~10%, respectively[6]. Securin is another IDP for which the original CU type description had to be changed to a MU.

The PreSMo concept seems duly acknowledged especially in recent years with many reports on the presence of PreSMos in free MU-type IDPs[6]. Even though the potential formation of local structural order by the eIF4E-binding segment in 4EBP1 was predicted by what is known as MoRF[21] and a recent mutation study showed the functional significance of the helical propensity of a short eIF4E-binding peptide (residues 51-67) in 4EBP1[22] no quantitative characterization on the formation of the pre-structure helix *per se* by the eIF4E-binding segment in a full or in a sufficiently long 4EBP1 construct with several residues flanking the eIF4E-binding segment has been carried out. Two mechanistic models, conformational selection (CS) and induced fit are currently in use to describe protein-protein interactions. In the case of globular proteins some were found to follow the former mechanism while others the latter. Recent results indicated that IDP-target binding cannot be fully accounted for only by the coil → helix IF mechanism[23,24]. Yet the fact that the IF has been considered predominantly for IDPs can probably be ascribed to the early view that IDPs were entirely unstructured. While there are at least a few dozen cases of the PreSMo structures known in free IDPs the cases where conformations of PreSMos both in the free and the target-bound state are very rare; examples are the p53 TAD helix and mdm2[3,25], the two turn motifs of p53 TAD and RPA[26], the turn II PreSMo of p53 TAD and p62[27] and the KID-KIX pair[10]. Our result on the structure of the free eIF4E-binding PreSMo along with its previously known conformation in its eIF4E-bound state, formed by exactly the same residues, is a meaningful addition to the above list. It suggests that eIF4E-4EBP1 binding may follow an initial conformational selection of the helix PreSMo in 4EBP1 by eIF4E followed by further structural induction into a more stable helix. Here, we underline again that presence of a PreSMo itself is not an evidence for conformational selection and that accurate determination of the IDP-target binding mechanism requires much more work, e.g., binding kinetics measurement with PreSMo segment mutations, NMR relaxation dispersion experiments etc. Nevertheless, we anticipate that this report contribute to the shift of our view on the IDP-target binding mechanism from the predominant IF to a combination of CS and IF since the CU nature of the full-length eIF4E-free 4EBP1 that played an important role in the conception of the coil→helix IF proposal along with the misleading original report on the KID-KIX binding[10] is now revised. It was probably the rarity of such data that did not allow one to seriously consider the conformational selection of a PreSMo by a target as an alternative IDP-target binding mechanism. In retrospect, the coil → helix IF mechanism for IDP-target binding was based only on a very limited number of NMR data and appears to have been generalized without thorough verification on a statistically significant number of systems[9,28].

The authors wish to thank J. A. Ferretti, and J. J. Han, for carefully reading the manuscript. This work was supported by UGM0021011 from Korea Research Council of Fundamental Sciences and Technology (KRCF) and a collaborative research project (C11005) (to K.H.) and the Max Planck Society and the EU (ERC grant agreement number 233227) (to C.G.). The computing resources were supported by the strategic support program (KCS-2011-C2-15) of Korea Institute of Science and Technology Information (KISTI).

Notes and references

[a] *Biomedical Translational Research Center, Division of Convergent Biomedical Research, Korea Research Institute of Bioscience and Biotechnology (KRIBB), 125 Gwahak-ro, Yuseong-gu, Daejeon, 305-806 Korea. Fax: +82 42 860 4259; Tel:+82 42 860 4250; E-mail:khhan600@kribb.re.kr*
[b] *Department of Bioinformatics, University of Science and Technology (UST), 113 Gwahak-ro, Yuseong-gu, Daejeon, 305-333 Korea.*
[c] *Department of NMR-based Structural Biology, Max-Planck Institute for Biophysical Chemistry, Göttingen, Germany*
‡These suthors contributed equally to this work.
† Electronic Supplementary Information (ESI) available: Experimental details, supplimentary tables, and figures. See DOI: 10.1039/b000000x/

1 V. N. Uversky and A. K. Dunker, *Biochim. Biophys. Acta,* 2010, **1804**, 1231-1264.
2 J. Marcotrigiano, A.-C. Gingras, N. Sonenberg, and S. K. Burley, *Mol. Cell,* 1999, **3**, 707-716.
3 H. Lee, K. H. Mok, R, Muhandiram, K.-H. Park, J.-E. Suk, D.-.H. Kim, J. Chang, Y. C. Sung, K. Y. Choi, and K.-H. Han, *J. Biol. Chem.,* 2000, **275**, 29426-29436.
4 T. Bartels, J. G. Choi, and D. J. Selkoe, *Nature,* 2011, **477**, 107-110.
5 D.-H. Kim, Y. Ni, S.-H. Lee, S. Urban, and K.-H. Han, *BMB rep.,* 2008, **41**, 640-644.
6 S.-H. Lee, D.-H. Kim, J. J. Han, E.-J. Cha, J.-E. Lim, Y.-J. Cho, C. Lee, and K.-H. Han, *Curr. Protein Pept. Sci.,* 2012, **13**, 34-54.
7 R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu, *Chem. Rev.*, 2014, **114**, 6589-6631.
8 C. M. Fletcher and G. Wagner, *Protein Sci.,* 1998, **7**, 1639-1642.
9 A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswood, W. Chiu, E. C. Garner, and Z. Obradovic, *J. Mol. Graph. Model.,* 2001, **19**, 26-59.
10 I. Radhakrishnan, G. C. Perez-Alvarado, D. Parker, H. J. Dyson, M. R. Montminy, and P. E. Wright, *Cell,* 1997, **91**, 741-752.
11 M. Uesugi, O. Nyanguile, H. Lu, A. J. Levine, and G. L. Verdine, *Science,* 1997, **277**, 1310-1313.
12 E. Giniger and M. Ptashne, *Nature,* 1987, **330,** 670–673..
13 P. B. Sigler, *Nature,* 1988, **333**, 210-212.
14 R. L. Baldwin and B. H. Zimm, *Natl. Acad. Sci. U. S. A.,*2000, **97**, 12391-12392.

15 D. Neri, M. Billeter, G. Wider, and K. Wuthrich, *Science,* 1992, **257**, 1559-1563.
16. Q.-X. Hua, W.-H. Jia, B. P. Bullock, J. F. Habener, and M. A. Weiss, *Biochemistry*, 1988, **37**, 5858.
17 V. Iešmantavičius, J. Dogan, P. Jemth, K. Teilum, and M. Kjaergaard, *Angew. Chem. Int. Ed*., 2014, **53**, 1548-1551.
18 R. Schneider, J.-R.Huang, M. Yao, G. Communie, V. Ozenne, L. Mollica, L. Salmon, R. M. Jensen, and M. Blackledge, *Mol.BioSyst.,* 2012, **8**, 58-68.
19 K. R. Shoemaker, P. S. Kim, E. J. York, J. M. Stewart, and R. L. Baldwin, *Nature,* 1987, **326**, 563-567.
20 I. Radhakrishnan, G. C. Perez-Alvarado, H. J. Dyson, and P. E. Wright, *FEBS Lett*.,1998, **430**, 317-322.
21 A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, A. K. Dunker, and V. N. Uversky, *J. Mol, Biol.*, 2006, **362**, 1043-59.
22 S. Tait, K. Dutta, D. Cowburn, J. Warwicker, A. J. Doig, and J. E. G. McCarthy, *Proc. Natl, Acad, Sci. U. S. A.,* 2010, **107**, 17627-17632.
23 J. M. R. Baker, R. P. Hudson, V. Kanelis, W.-Y. Choy, P. H. Thibodeau, P. J. Thomas, and J. D. Forman-Kay, *Nat. Struct. Mol. Biol*., 2007, **14**, 738-745.
24 D. D. Boehr, R. Nussinov, and P. E. Wright, *Nat. Chem. Biol*. 2009, **5**, 789-796.
25 P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine, and N. P. Pavletich, *Science,* 1996, **274**, 948-953
26 E, Bochkareva, L. Kaustov, A. Ayed, G.-S. Yi, A. Pineda-Lucena, J. C. C. Liao, A. L. Okorokov, J. Milner, C. H. Arrowamith, and A. Bochlarev, *Proc. Natl. Acad. Sci. U. S. A.,*2005, **102**, 15412-15417.
27 P. Di Lello, L. M. M. Jenkins, T. N. Jones, B. D. Nguyen, T. Hara, H. Yamaguchi, P. Legault, and J. G. Omichinski, *Mol. Cell*. 2006, **22**, 731-740.
28 T. Chouard, *Nature,* 2011, **471**, 151-153.

