Accepted Article

Title: The dynamic multisite interactions between two intrinsically disordered proteins

Authors: Shaowen Wu, Dongdong Wang, Jin Liu, Yitao Feng, Jingwei Weng, Yu Li, Xin Gao, Jianwei Liu, and Wenning Wang

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

To be cited as: *Angew. Chem. Int. Ed.* 10.1002/anie.201701883
Angew. Chem. 10.1002/ange.201701883

Link to VoR: <http://dx.doi.org/10.1002/anie.201701883>
<http://dx.doi.org/10.1002/ange.201701883>

The dynamic multisite interactions between two intrinsically disordered proteins

Shaowen Wu,^[1,3] Dongdong Wang,^[1,3] Jin Liu,^[1,3] Yitao Feng,^[1] Jingwei Weng,^{*[1]} Yu Li,^[2] Xin Gao,^[2] Jianwei Liu,^{*[1]} and Wenning Wang^{*[1]}

Abstract: Protein interactions involving intrinsically disordered proteins (IDPs) comprise a variety of binding modes, from the well characterized folding upon binding to dynamic fuzzy complex. To date, most studies concern the binding of an IDP to a structured protein, while the interaction between two IDPs is poorly understood. In this study, we combined NMR, smFRET, and molecular dynamics (MD) simulation to characterize the interaction between two IDPs, the C-terminal domain (CTD) of protein 4.1G and the nuclear mitotic apparatus (NuMA) protein. It is revealed that CTD and NuMA form a fuzzy complex with remaining structural disorder. Multiple binding sites on both proteins were identified by MD and mutagenesis studies. Our study provides an atomic scenario in which two IDPs bearing multiple binding sites interact with each other in dynamic equilibrium. The combined approach employed here could be widely applicable for investigating IDPs and their dynamic interactions.

Protein 4.1 is a ubiquitously expressed adaptor protein that cross-links spectrin/actin cytoskeleton to a wide variety of membrane proteins.^[1] In vertebrates, protein 4.1 family has four members: 4.1R, 4.1G, 4.1N, and 4.1B. Early studies have reported that 4.1R/4.1N interacts with NuMA and regulates the assembly of the mitotic spindle.^[2] Recently, 4.1R/4.1G-NuMA interaction was found to play a major role in the anaphase-specific localization of NuMA in symmetrically dividing Hela cells,^[3] and in the maintenance of NuMA cortical stability and spindle orientation in asymmetrically dividing keratinocytes.^[4] Protein 4.1 interacts with NuMA's C-terminal region through its highly conserved C-terminal domain (CTD) (Figure 1a).^[2-5] CTD is unique for the protein 4.1 family members, and is conserved through invertebrates to vertebrates.^[1] It was first defined as a 22/24 kDa domain by its resistance to chymotryptic digestion of 4.1R.^[6] More detailed biochemical characterizations of CTD showed that a 15-kDa fragment is stable under chymotryptic digestion and is most likely a globular folded domain.^[7] Recent NMR characterizations of 4.1R-CTD demonstrated that it is most likely intrinsically disordered or unstructured.^[8] On the other

hand, the C-terminal region of NuMA, which contains binding sites for several proteins involved in the control of mitosis,^[2a, 9] is also unstructured according to sequence analysis and previous structural studies.^[10] A conventional diagram of IDP interaction involves a disorder-to-order transition to form a stable structured complex. Alternatively, recent studies have shown that IDPs could also form dynamic complex with remaining structural disorder. However, in most of the cases the binding partner of IDP is a stably folded protein. It is intriguing that how two IDPs form specific binding, and what kind of binding model they follow. A fully quantitative characterization of intrinsically disordered proteins (IDPs) and their interactions is a big challenge. Each of the current techniques could only provide limited information on the heterogeneous conformational states of IDPs. A potentially applicable approach is combining complementary techniques, especially experimental and computational ones to decipher the broad structure ensemble of IDPs and their interactions. In this work, we combine NMR, MD simulation, and smFRET to characterize the interaction between 4.1G-CTD and NuMA in atomic detail. It is shown that 4.1G-CTD and NuMA form a dynamic complex with both proteins unstructured. The dynamic multisite interaction mode is characterized in atomic detail. We started by verifying the direct interaction between the CTD of 4.1G (aa. 886-1005) and the C-terminal region of NuMA (aa. 1788-1832) through GST pull-down (Figure 1b). To find the optimal binding regions, we purified various fragments of the two proteins and characterized the binding affinities by isothermal titration calorimetry (ITC) measurements (Figure 1c). It turns out that 4.1G886-1005 is sufficient for binding to NuMA1800-1825, and the C-terminal half of 4.1G (4.1G939-1005) is the major binding region (Figure 1c,d).

Figure 1. Biochemical characterizations of 4.1G-NuMA interaction. (a) The domain organization of 4.1G and NuMA. The 4.1G binding motif (aa. 1800-1825) in NuMA CTD is indicated. (b) GST pull-down assay showing the

- [1] S.Wu, D. Wang, J. Liu, Y. Feng, J. Weng, J. Liu, W. Wang Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Department of Chemistry, and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433 E-mail: jweng@fudan.edu.cn for Dr. J. Weng, jianweiliu@fudan.edu.cn for Dr. Jianwei Liu, wnwang@fudan.edu.cn for Prof. W. Wang
- [2] Y. Li, Prof. X. Gao King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center (CBRC), Computer, Electrical and Mathematical Sciences and Engineering (CEMSE), Division, Thuwal, 23955, Saudi Arabia E-mail: xin.gao@kaust.edu.sa for Prof. X. Gao
- [3] These authors contributed equally to this work.

interaction between GST-4.1G-CTD and NuMA 1788-1832. (c) The binding affinity of NuMA 1800-1825 to various 4.1G segments at the CTD region measured by ITC. (d) ITC experiments of the interactions of NuMA with 4.1G868-1005 and 4.1G939-1005.

To explore the structural basis of the interaction between CTD and NuMA, we used NMR spectroscopy to characterize the 4.1G868-1005 fragment. The ^1H - ^{15}N HSQC spectrum of 4.1G868-1005 displays limited amide proton chemical shift dispersion (Figure S1), a diagnostic characteristic of intrinsically disordered proteins. We failed in backbone assignment of 4.1G868-1005 due to protein sample aggregation. Since the shorter fragment 4.1G939-1005 exhibits comparable affinity for NuMA1800-1825 (Figure 1c), we characterized this protein in details. The ^1H - ^{15}N HSQC spectrum of 4.1G939-1005 also displays limited chemical shift dispersion, and ~ 60% backbone assignments were achieved (Figure 2a). The secondary structure propensity (SSP) values^[11] were determined using chemical shifts. The SSP values demonstrate that although most regions do not form regular secondary structures, a stretch of residues between Gln970 and Arg980 has up to 70% weighted likelihood to form α -helix (referred to αA hereafter, Figure 2b). This observation is in agreement with the sequence-based disorder prediction using PONDR (Figure S2).

Figure 2. 4.1G CTD region is intrinsically disordered. (a) The ^1H - ^{15}N HSQC spectrum of 4.1G939-1005 at pH 6.8 shows that the protein is intrinsically disordered. (b) Secondary structure estimated from the $\text{C}\alpha$ and $\text{C}\beta$ chemical

shifts of 4.1G939-1005. (c) The agreement between the experimental $\text{C}\alpha$ chemical shifts (left) and secondary chemical shifts (right) of 4.1G939-1005 with those predicted from the structure ensemble of REMD simulation. (d) The representative structures of the first 15 clusters of the structure ensemble of 4.1G939-1005 from the REMD simulation.

To obtain the structural ensemble of 4.1G939-1005, we performed a total of 9.6 μs replica-exchange molecular dynamics (REMD) simulation (see Methods for more details). The chemical shifts and secondary chemical shifts (SCSSs) of $\text{C}\alpha$ and $\text{C}\beta$ predicted from the MD-derived structure ensemble are in good agreement with the NMR measurements (Figure 2c and Figure S3). The agreement here is comparable and even better than other simulation works.^[12] RMSD-based cluster analysis provides us a clearer picture of the structural ensemble. The top 15 clusters of the conformational ensemble exhibits remarkable structural heterogeneity (Figure 2d). Nevertheless, consistent with the NMR experiment, the αA helix is remarkably stable in almost all the clusters (Figure 2d). REMD simulation also demonstrates that 4.1G-CTD is generally globular. It was proposed that the low hydrophobicity and a high net charge is an important prerequisite for the absence of a compact structure.^[13] By calculating the charge/hydrophathy (C/H) ratios of 4.1G-CTD fragments, we found that 4.1G-CTD is on the edge of being natively unfolded and folded globular proteins (Figure S4). This explains why the previous biochemical study has identified CTD as a globular domain.^[7] In the broad spectrum of IDPs, 4.1G-CTD apparently belongs to the partially or weakly folded proteins rather than the extended coils.

To characterize the NuMA-4.1G interaction interface, we performed NMR titration experiments. Titration of unlabeled NuMA1800-1825 step-by-step into 4.1G939-1005 leads to resonance line broadening for most of the peaks on the ^1H - ^{15}N HSQC spectrum (Figure S5). This phenomenon could result from the increased effective correlation time, restricted conformational motion, and/or exchange between the free and bound states on the microsecond to millisecond time scales. All the peaks did not show obvious chemical shift changes during titration (Figure S5), suggesting that 4.1G939-1005 remained structurally disordered upon NuMA binding. In the cases where IDPs fold into a stable structure upon complex formation, such as the case of pKID/KIX complex,^[14] new NMR peaks appear at chemical shifts characterizing structured proteins upon complex formation. There is no such evidence in our system. It is most likely that 4.1G-CTD does not undergo an overall disorder-to-order transition upon NuMA binding and the two proteins form a dynamic complex without a static conformation. We also produced a fusion protein by covalently linking NuMA1800-1825 to the C-terminus of 4.1G939-1005. The HSQC spectrum of this fusion protein also demonstrates characteristics of IDP with many broadened signals (Figure S6), supporting the above conclusion.

In order to obtain atomic details of the interaction, we performed REMD simulation for the 4.1G939-1005/NuMA1800-1825 complex (see Methods for more details). As a control, we also performed REMD simulation of the isolated NuMA1800-1825 peptide, which exhibits conformational heterogeneity in solution (data not shown). The simulation results of 4.1G939-1005/NuMA1800-1825 demonstrate that complex formation

does not stabilize the conformation of CTD or NuMA, both of which exhibit high structural plasticity (Figure 3a). The N-terminal part of NuMA peptide tends to form a short α -helix, but both secondary structure and overall conformation of NuMA peptide are highly variable among the top 15 clusters (Figure 3a). The structure features of 4.1G939-1005 in the complex share similarities with those of the free form 4.1G939-1005. Secondary structure analysis showed that the content of α -helix increases slightly with respect to the free form 4.1G, while the variation of β -strand contents is less obvious (Figure S7). The α A helix is still very stable in the complex (Figure S7).

Figure 3. The structure basis of the interaction between 4.1G939-1005 and NuMA1800-1825. (a) The representative structures of the first 15 clusters of the structure ensemble of the 4.1G939-1005/NuMA1800-1825 complex from the REMD simulation. NuMA peptide is colored orange. (b) The contact map between 4.1G939-1005 and NuMA 1800-1825 based on the REMD simulation trajectory. (c) Mutations on 4.1G or NuMA designed based on the contact map to attenuate or disrupt the interaction between the two proteins.

The overall conformation of 4.1G939-1005 is obviously affected by NuMA binding. We compared the structures of 4.1G939-1005 in all the 98 clusters of 4.1G/NuMA complex with those of the isolated 4.1G939-1005 and listed the most similar counterparts with minimal C α RMSDs in Table S1. It turns out that the RMSDs between these counterpart pairs are all larger than 0.4 nm (Table S1), indicating that NuMA binding modulates the structure ensemble of 4.1G939-1005 remarkably. The structure ensemble clearly shows that the interaction between NuMA and

4.1G939-1005 is highly variable with multiple binding sites (Figure 3a). This is consistent with the NMR observation that most of the peaks are broadened upon NuMA binding (Figure S5). We calculated the residue specific contact probabilities between 4.1G939-1005 and NuMA based on the simulation trajectories (Figure 3b). The contact map clearly demonstrates the multivalent nature of the interaction. Both proteins bear multiple binding sites and interact with each other in dynamic equilibrium. However, a few binding sites show remarkably high contact probabilities (Figure 3b), which implicate binding specificity. One frequently contacting region results from the hydrophobic interactions between the 988VTRVVV993 motif on 4.1G939-1005 and the 1814IINITM1819 motif on NuMA. In addition, the negatively charged residues at the C-terminal region of 4.1G939-1005 interact with the positively charged residues at the N-terminal region of the NuMA peptide. The positively charged N-terminus of NuMA also interacts with a negatively charged motif between Asp9634.1G and Asp9694.1G. To verify the binding site predictions from the REMD simulation, we generated several mutations at the sites with high contact probabilities on both 4.1G and NuMA and measured the binding affinities of the mutant proteins. All the mutations attenuated or disrupted the binding between 4.1G and NuMA (Figure 3c). Therefore, the structure ensemble correctly reflects the binding mode between 4.1G939-1005 and NuMA peptide. Multivalent dynamic binding has been revealed previously in the interaction of IDPs with structured proteins, such as the case of pSic1/CDC4.^[15] The multivalent dynamic binding mode in our system is more complicated since the two binding partners are simultaneously undergoing significant conformational fluctuations.

The severe line broadening on the NMR spectrum prevents us from further exploring the structural and dynamics information using NMR spectroscopy. Therefore, we performed smFRET measurements for 4.1G939-1005 and its complex with NuMA. We generated two pairs of labelling sites for FRET donor and acceptor (939/982 and 982/1005) by mutating Thr939, Gln982, and Asp1005 to cysteine, respectively (Figure 4a). In both labelling systems (referred to as 4.1G (N) and 4.1G (C), respectively), the FRET efficiency distributions are broad and apparently could not be described as a single Gaussian peak (Figure S8&S9), suggesting complex conformational heterogeneity. The shot noise contributed a small fraction of the width of FRET histogram (Table S2 and see methods in SI). The donor-acceptor orientation, photophysics processing of dye may also introduce the broadening of FRET histogram that is common in the smFRET measurement. However, the significant change of histogram profile upon NuMA binding and their non-uniform shape demonstrated the presence of significant conformational heterogeneity of 4.1G. To compare smFRET measurement and MD simulation directly, the FRET efficiency was converted to inter-dye distance and the donor-acceptor distances of the MD snapshots were calculated using available volume method (see Methods).^[16] The distance distribution profiles and their variation shifts upon NuMA binding are consistent between experiment and simulation (Figure 4b&c). In both labelling systems, the inter-dye distances increase upon complex formation. Notably, the distance distributions of smFRET measurement has a systematic deviation from those of

the MD simulations (Figure 4b&c). Such discrepancy could be caused by the errors from both experiment and simulation (see Methods for more details). However, the qualitative agreement cross-validates the structure ensembles derived from REMD simulations.

Figure 4. smFRET experiments of 4.1G939-1005 and 4.1G939-1005/NuMA. (a) Diagram showing the positions of labeled fluorescence donor and acceptor on 4.1G939-1005. The donor-acceptor pair on the left is 939-982 and on the right is 982-1005. (b) Inter-dye distance distributions converted from smFRET efficiency for the two labelling schemes of both 4.1G and 4.1G/NuMA complex. (c) Distributions of the inter-dye distances obtained from the structure ensemble of REMD simulations.

Examination of the smFRET traces of both 4.1G (N) and 4.1G (C) systems reveals stochastic transitions between various FRET efficiencies (Figure S10a-d). This stochastic switching behaviour was also observed in other compact globular IDPs distinct from the extended random coils^[17] and is remarkably different from conventional structured proteins (Figure S10e). We performed hidden Markov model (HMM) analyses of the trajectories^[18] (see Methods for more details), and the transition density plots are shown in Figure 5. Both 4.1G(N) and 4.1G(C) undergo conformational transitions between many discrete states. Notably, NuMA binding changes the pattern of the transition density maps, but the complex systems still fluctuate among many states (Figure 5), indicating that the complex remains the dynamic and unstructured nature of the isolated 4.1G939-1005. This confirms the conclusion derived from the above NMR experiment and MD simulations. Overall, the smFRET efficiency distributions cross-validate the MD-derived structure ensembles, and the HMM analysis of smFRET trajectories confirms that the 4.1G-CTD and NuMA form a dynamic complex with similar conformational fluctuations with that of the isolated 4.1G939-1005.

Figure 5. Conformational fluctuations revealed by transition density plots (TDPs) based on HMM analysis. TDP of (a) 4.1G(N). (b) 4.1G(N)/NuMA. (c) 4.1G(C). (d) 4.1G(C)/NuMA...

In summary, we combined NMR, MD simulation and smFRET methods to characterize the interaction between 4.1G CTD and NuMA. It has been revealed that the two IDPs form a dynamic fuzzy complex^[19] with both proteins remaining disordered. Both proteins bearing multiple binding sites interact with each other in dynamic equilibrium. This extremely dynamic interaction mode between IDPs has been implicated previously in a few studies,^[20] but to our knowledge, this study characterizes the detailed molecular mechanism for the first time.

The dynamic nature of the complex may facilitate the modulation of 4.1G-CTD/NuMA interaction in the functional context. It has been reported that the cortical localization of NuMA during anaphase is governed by interaction with phosphoinositides (PIs) at a region that encompasses the 4.1G binding site. The fuzziness of 4.1G/NuMA interaction may ease the competitive binding of PIs to NuMA. Except for NuMA, many binding partners of 4.1-CTD are C-terminal tails of membrane proteins, which are mostly disordered or unstructured. Therefore forming dynamic fuzzy complex like the case of NuMA may be a common theme of 4.1-CTD mediated protein-protein interactions. These highly dynamic protein-protein interactions of IDPs are quite challenging for conventional biophysical techniques, thereby the combined approach employed in this study could be an effective way in studying IDPs and their dynamic fuzzy complexes.

Acknowledgements

This work was supported by National Major Basic Research Program of China (2016YFA0501702), National Science Foundation of China (21473034, 21375028, 21273188), Specialized Research Fund for the Doctoral Program of Higher Education (20130071140004). X. G. and Y. L. thanks the

support by funding from King Abdullah University of Science and Technology (KAUST). This research made use of the resources of the computer clusters at KAUST and the computer center at Fudan University.

Keywords: Intrinsically disordered protein • CTD of protein 4.1 • NMR • smFRET • MD simulation

- [1] a) K. B. Hoover, P. J. Bryant, *Curr. Opin. Cell Biol.* 2000, 12, 229-234; b) A. J. Baines, H. C. Lu, P. M. Bennett, *Biochim. Biophys. Acta* 2014, 1838, 605-619.
- [2] a) S. N. Mattagajasingh, S. C. Huang, J. S. Hartenstein, M. Snyder, V. T. Marchesi, E. J. Benz, *J. Cell Biol.* 1999, 145, 29-43; b) K. Ye, D. A. Compton, M. M. Lai, L. D. Walensky, S. H. Snyder, *J. Neurosci.* 1999, 19, 10747-10756.
- [3] T. Kiyomitsu, I. M. Cheeseman, *Cell* 2013, 154, 391-402.
- [4] L. Seldin, N. D. Poulson, H. P. Foote, T. Lechler, *Mol. Biol. Cell* 2013, 24, 3651-3662.
- [5] S. N. Mattagajasingh, S. C. Huang, E. J. Benz Jr, *Clin. Transl. Sci.* 2009, 2, 102-111.
- [6] T. L. Leto, V. T. Marchesi, *J. Biol. Chem.* 1984, 259, 4603-4608.
- [7] C. Scott, G. W. Phillips, A. J. Baines, *Eur. J. Biochem.* 2001, 268, 3709-3717.
- [8] M. A. Trevino, M. Rodriguez-Rodriguez, I. Correias, M. Marcilla, J. P. Albar, M. Rico, M. A. Jimenez, M. Bruix, *BMC Biochem.* 2010, 11, 7.
- [9] a) Q. Du, P. T. Stukenberg, I. G. Macara, *Nat Cell Biol* 2001, 3, 1069-1075; b) L. Haren, A. Merdes, *J. Cell Sci.* 2002, 115, 1815-1824.
- [10] J. Zhu, W. Wen, Z. Zheng, Y. Shang, Z. Wei, Z. Xiao, Z. Pan, Q. Du, W. Wang, M. Zhang, *Mol Cell* 2011, 43, 418-431.
- [11] J. A. Marsh, V. K. Singh, Z. Jia, J. D. Forman-Kay, *Protein Sci.* 2006, 15, 2795-2804.
- [12] a) Y. Wang, X. K. Chu, S. Longhi, P. Roche, W. Han, E. K. Wang, J. Wang, *Proc. Natl. Acad. Sci. U. S. A.* 2013, 110, E3743-E3752; b) Y. Luo, B. Ma, R. Nussinov, G. Wei, *J Phys Chem Lett* 2014, 5, 3026-3031.
- [13] V. N. Uversky, J. R. Gillespie, A. L. Fink, *Proteins* 2000, 41, 415-427.
- [14] K. Sugase, H. J. Dyson, P. E. Wright, *Nature* 2007, 447, 1021-1025.
- [15] T. Mittag, S. Orlicky, W. Y. Choy, X. Tang, H. Lin, F. Sicheri, L. E. Kay, M. Tyers, J. D. Forman-Kay, *Proc Natl Acad Sci U S A* 2008, 105, 17772-17777.
- [16] a) S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Muller, C. A. Seidel, *J Am Chem Soc* 2011, 133, 2463-2480; b) K. Walczewska-Szewc, B. Corry, *Phys Chem Chem Phys* 2014, 16, 12317-12326.
- [17] U. B. Choi, J. J. McCann, K. R. Weninger, M. E. Bowen, *Structure* 2011, 19, 566-576.
- [18] J. W. van de Meent, J. E. Bronson, C. H. Wiggins, R. L. Gonzalez, Jr., *Biophys J* 2014, 106, 1327-1337.
- [19] P. Tompa, M. Fuxreiter, *Trends Biochem.Sci.* 2008, 33, 2-8.
- [20] a) A. Sigalov, D. Aivazian, L. Stern, *Biochemistry* 2004, 43, 2049-2061; b) A. B. Sigalov, A. V. Zhuravleva, V. Y. Orekhov, *Biochimie* 2007, 89, 419-421; c) M. S. Pometun, E. Y. Chekmenev, R. J. Wittebort, *J. Biol. Chem.* 2004, 279, 7982-7987.

Entry for the Table of Contents (Please choose one layout)

Layout 1:

COMMUNICATION

Text for Table of Contents

Dynamic multisite interactions between two IDPs: Combined NMR, MD simulation, and smFRET studies reveal that two intrinsically disordered proteins, 4.1G-CTD and NuMA, form dynamic fuzzy complex with both proteins remaining disordered. Two proteins bearing multiple binding sites interact with each other in dynamic equilibrium.

S. Wu, D. Wang, J. Liu, Y. Feng, J. Weng*, Y. Li, X. Gao, J. Liu*, W. Wang*

Page No. – Page No.

The intrinsically disordered C-terminal domain of protein 4.1G and NuMA form a dynamic fuzzy complex

Layout 2:

COMMUNICATION

((Insert TOC Graphic here))

Author(s), Corresponding Author(s)*

Page No. – Page No.

Title

Text for Table of Contents