JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

A Conformational Ensemble Derived Using NMR Methyl Chemical Shifts Reveals a Mechanical Clamping Transition That Gates the Binding of the HU Protein to DNA

Arvind Kannan, Carlo Camilloni, Aleksandr B. Sahakyan, Andrea Cavalli, and Michele Vendruscolo*

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

Supporting Information

ABSTRACT: Recent improvements in the accuracy of structure-based methods for the prediction of nuclear magnetic resonance chemical shifts have inspired numerous approaches for determining the secondary and tertiary structures of proteins. Such advances also suggest the possibility of using chemical shifts to characterize the conformational fluctuations of these molecules. Here we describe a method of using methyl chemical shifts as restraints in replica-averaged molecular dynamics (MD) simulations, which enables us to determine the conformational ensemble of the HU dimer and characterize the range of motions accessible to its flexible β-arms. Our analysis suggests that the bending action of HU on DNA is mediated by a mechanical clamping mechanism, in which metastable structural intermediates sampled during the hinge motions of the β-arms in the free state are presculpted to bind DNA. These results illustrate that using side-chain chemical shift data in conjunction with MD simulations can provide quantitative information about the free energy landscapes of proteins and yield detailed insights into their functional mechanisms.

A wide range of proteins populate conformationally highly heterogeneous states whose structures are not readily amenable to X-ray crystallography. Quantitative information about the structural properties of such states can, at least in principle, be extracted from NMR chemical shifts, which have shown promise for protein structure determination.[1] Since chemical shifts are the NMR observables that can be measured under the widest range of conditions and with the greatest accuracy, they constitute an invaluable experimental tool for probing the structure and dynamics of challenging biomolecular systems. Considerable advances have been made recently in the development of methods for predicting protein dihedral angles, secondary structure populations, and tertiary structures from chemical shifts.[1,2] Moreover, it is now known that the accuracy of structure-based chemical shift predictions can improve significantly when the calculations are averaged over ensembles of protein conformations,[3] suggesting that chemical shift data can be used to generate accurate structural ensembles that reflect the dynamics of proteins in solution. We recently established the feasibility of this approach by implementing chemical shifts for backbone atoms as structural restraints in replica-averaged molecular dynamics (MD) simulations.[4]

Here we report a further step in this direction by demonstrating that the use of chemical shift restraints can be extended to side chains, leading to an increased consistency between the reconstructed conformational ensembles and experimental data. We illustrate this approach by applying it to a highly dynamical protein complex, the DNA-bending HU dimer. NMR studies of HU have revealed that this protein can exist in a multitude of conformations depending on the orientation of the flexible β-hairpin loops in its DNA-binding "arms".[5] Furthermore, relaxation experiments have identified long-time-scale hinge motions of the β-arms relative to the core, which are hypothesized to lower the kinetic barrier for DNA association and dissociation.[6] We demonstrate that MD simulations with methyl (Me) chemical shift restraints provide atomistic insights into the nature of the loop motions that complement those from previous experiments, suggesting a "mechanical clamping" mechanism by which HU can bind DNA.

In analogy to a recent method of using backbone chemical shifts as structural restraints,[4,7] we incorporated the experimental Me resonances into a molecular mechanics force field (E_{total}) by means of a restraint potential (E_{CS}),

$$E_{\text{total}} = E_{\text{MM}} + E_{\text{CS}} \tag{1}$$

where E_{MM} is a standard force field (see SI) and

$$E_{\text{CS}} = \sum_i \sum_j \kappa_{ij} (\overline{\delta}_{ij}^{\text{calc}} - \delta_{ij}^{\text{exp}})^2 \tag{2}$$

Here the index i runs over all Ala, Ile, Leu, Thr, and Val residues in the protein, while the index j runs over all nonequivalent Me protons in each residue. The weighting factors κ_{ij} set the relative strength of each restraint and are chosen to be as large as possible without compromising the stability of the trajectories.[8] Chemical shifts are calculated as averages over four replicas[9] to mimic the ensemble-averaging inherent in the experimental measurements[10] within the maximum entropy approach,[8,11]

$$\overline{\delta}_{ij}^{\text{calc}} = \frac{1}{4} \sum_{k=1}^{4} \delta_{\text{calc}}^{ij,k} \tag{3}$$

Methyl chemical shifts for each replica in eq 3 are calculated at every time step of the simulation using CH3Shift, which parametrizes the chemical shifts as differentiable functions of the atomic coordinates of the protein.[12] This feature of the

Received: October 14, 2013
Published: December 31, 2013

   dx.doi.org/10.1021/ja4105396 | J. Am. Chem. Soc. 2014, 136, 2204−2207

prediction engine allows for the calculation of pseudo-forces which, in conjunction with the force field, drive the system toward conformations that are consistent with the experimental measurements. Note that while an infinite number of conformational ensembles could in principle be consistent with the available NMR parameters, the maximum entropy approach assumes the equilibrium distribution of a given force field as a prior, and finds the specific ensemble corresponding to the minimal perturbation of the potential energy surface required for consistency with the experimental data.[8]

To validate the present approach, we ran a series of simulations of four representative proteins (Figure S1) involving (1) no restraints, (2) restraints on only backbone chemical shifts, (3) restraints on only Me chemical shifts, and (4) restraints on both backbone and Me chemical shifts. We simulated the unprocessed autophagy protein Atg8, the bone morphogenetic protein receptor BMPR-IA, and the G-protein regulator RGS10 alongside the HU dimer, so as to cover a range of protein folds and biological functionality. Since nearly complete backbone and Me resonance assignments are available for these systems,[13] they offer an informative test of the Me shift predictor through the investigation of all four restraint protocols in parallel. Furthermore, evidence from NMR, X-ray crystallography, and Raman spectroscopy suggests that all four proteins contain regions of high dynamic mobility that undergo conformational exchange on time scales ranging from nanoseconds to milliseconds,[6,14] making these systems ideal for validating our replica-averaged approach to structure calculations. The dynamics implied by these experiments stem from the presence of disordered termini in Atg8 and BMPR-IA, and from flexible loop regions in HU and RGS10.

We first assessed the ability of our Me chemical shift restraint protocol to generate structural ensembles that are consistent with the NMR spectra from which the restraints are derived. We back-calculated the backbone and side-chain chemical shifts from each trajectory using 4DSPOT,[3c] which parametrizes the chemical shifts as functions of time-averaged structural fluctuations in the protein, and compared these ensemble-averaged chemical shift predictions to the experimental values for all four ensembles (Figure 1). When the ensembles are generated using only Me chemical shift restraints, the agreement between the calculated and experimental backbone chemical shifts, which are not used as restraints, improves significantly (p = 0.003) relative to the unrestrained trajectory. Similar results are obtained when both the backbone and the Me chemical shifts are used as restraints, providing support for the present Me chemical shift restraint protocol, since the dependence of Me chemical shifts on local tertiary structure and side-chain packing has global effects that drive the backbone geometries toward their experimentally observed distributions. Conversely, the application of backbone chemical shift restraints alone also improves agreement with the Me chemical shift data, which are not used in the calculations, although to a lesser extent (p = 0.01). In addition, observed (see SI) that all three restrained ensembles (Me chemical shifts, backbone chemical shifts, and Me + backbone chemical shifts) exhibit some detectable improvement in the prediction of non-Me side-chain chemical shifts (C-*sc* and H-*sc*) that were not used as restraints (p = 0.04).

As an additional independent test of the quality of the ensembles, we computed other NMR parameters such as NOE distances and homonuclear three-bond $J(H^NH^\alpha)$ coupling constants from the trajectories and compared them to the experimental data where available (Figures 2 and S5 and Table S2). Statistically significant improvements were observed for all back-calculated NMR observables upon the application of side-chain restraints ($p < 10^{-7}$ for NOEs and < 0.01 for J-couplings). Furthermore, in Atg8 and BMPR-IA (the two systems exhibiting partial disorder), we compared time-averaged secondary structure populations from the trajectories with the predictions of the δ2D method[2b] to verify that the application of restraints corrects the tendency of the force field to provide relatively poor estimates of the α-helical and β-strand character of several residues (Figures S7 and S8; see the SI for a detailed discussion of the ensemble validation).

Having established the structural accuracy of the ensembles generated by the procedure described above, we focused our analysis on the HU dimer and asked whether the inclusion of Me chemical shifts is capable of uncovering new information about the conformational fluctuations of this protein complex. To

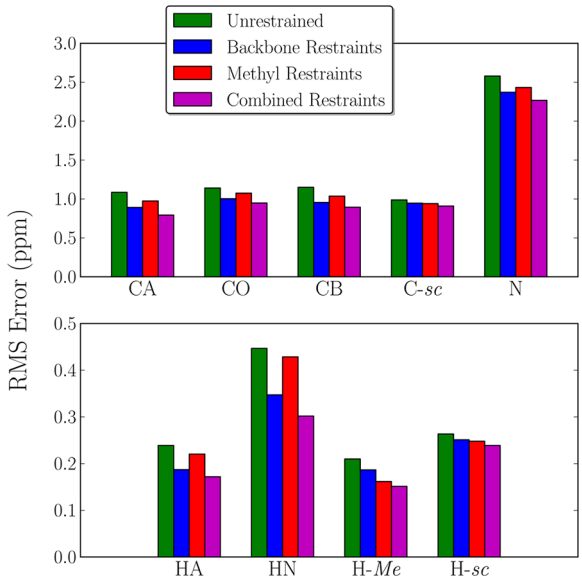

Figure 1. Comparison of the root-mean-square prediction errors from 4DSPOT for each atom type in the experimental data sets (BMRB 4503, BMRB 15730, BMRB 5097, and BMRB 4047) across the four simulated ensembles. Results are averaged over all four proteins. H-*Me* refers to methyl proton chemical shifts, while C-*sc* and H-*sc* refer to all other side-chain chemical shifts.

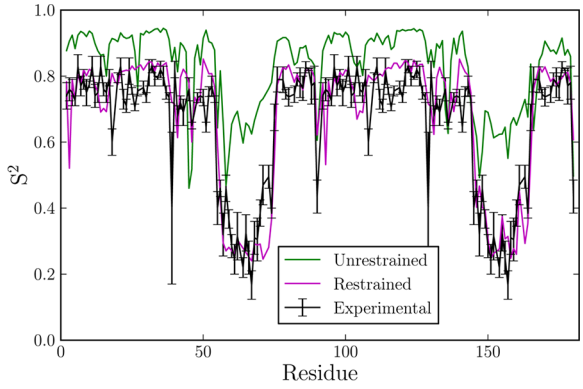

Figure 2. Comparison of S^2 order parameters for backbone N−H bonds in the HU dimer as derived from the unrestrained and restrained ensembles with the results of multiple-field ^{15}N cross-relaxation experiments.[6] Residue numbering begins at 1 and 91 for the two chains, respectively.

dx.doi.org/10.1021/ja4105396 | J. Am. Chem. Soc. 2014, 136, 2204−2207

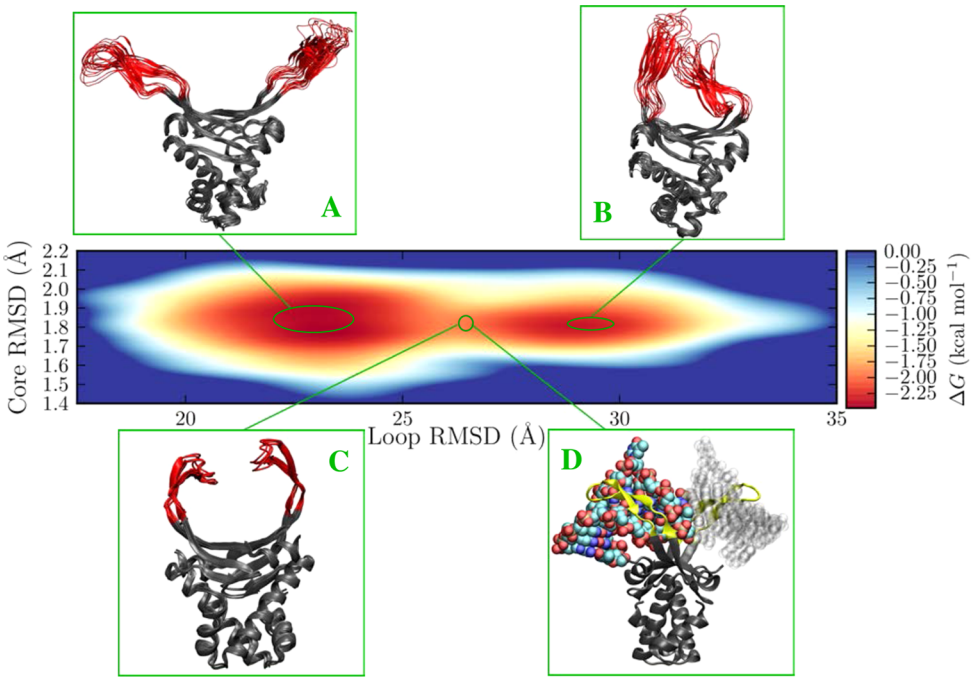

Figure 3. Free energy landscape of the HU dimer in the backbone−Me ensemble. RMSD values are reported relative to the starting structure in the unrestrained simulation (see SI for further details). Insets: (A) representative structures in the extended substate, (B) representative structures in the collapsed substate, (C) overlay of five intermediate loop conformers with the lowest RMSDs from the DNA-bound crystal structures of HU (PDB 1P71, 1P78, and 1P51), and (D) bound DNA fragment from PDB 1P71 superimposed on the lowest-energy structure from inset C (the structure is rotated by 90° relative to insets A−C).

represent dynamics in this system, we constructed a free energy landscape from the backbone−Me ensemble.

As shown in Figure 3, the free energy landscape of HU as a function of the core and loop RMSD can be clustered into two substates in which the loops either extend fully outward in V-shaped structures (inset A) or adopt collapsed configurations with the two chains in a triangular geometry (inset B). The unrestrained ensemble forces the loops into a single rigid conformation that is inconsistent with both of these substates and with the available dynamic data on the protein (Table S1c and Figures 2, S5, S11), reflecting the inadequacies of the force field that we used (see SI) in describing this system. The addition of restraints improves the sampling and allows the system to find its two free energy minima and undergo conformational exchange between them. The restrained ensemble captures well the dynamics of the unbound dimer in solution, as evidenced by the quantitative agreement between back-calculated and experimental S^2 order parameters (Figure 2).[6]

The atomistic resolution of the restrained HU trajectories clarifies the previously predicted hinge motion of the β-arms as a conformational exchange process between particularly stable extended and collapsed loop structures (insets A and B). We also observed structural intermediates in the transition between these two states, in which the arms have begun to collapse inward but have not yet locked together (inset C and Figure S10). The DNA-bound conformation of the dimer is sampled as one of these "intermediate" structures in the exchange process (inset D), and the bound-like conformers are superimposable with DNA-bound crystal structures of HU (backbone RMSD < 0.6 Å).[15]

Taken together with the high mobility and structural plasticity of the β-arms, the presence of DNA-bound loop conformers in the unbound ensemble suggests that interactions between HU and DNA proceed via a mechanism in which the normally metastable bound orientation of the arms is stabilized upon DNA binding. These results suggest that HU operates as a biomolecular clamp that freely samples nearby DNA segments in its extended state and then traps them in negatively supercoiled structures upon transitioning to its collapsed state. According to this model, DNA binding incurs a large entropic penalty because it requires a transition from a wide conformational basin in the free energy landscape to a much more restricted basin (Figure S9). Nevertheless, the ease with which a clamping mechanism can explore different substrate configurations may help explain HU's ubiquitous role as a sculptor of DNA architecture in many diverse biological systems.

To better understand how the observed hinge motion of the arms facilitates HU's interactions with DNA, we decomposed the loop dynamics into two principal components: a "clamp" axis and a "twist" axis (Figure 4A). The clamping motion gates the exchange process between the extended and collapsed substates and controls the amount of volume accessible to DNA. The twisting motion controls the ability of the arms to wrap around the minor groove of DNA, and the projected distance between the two hairpin tips along this axis is a measure of the arm curvature. Motions along these two dimensions are highly correlated (Figure 4B). Indeed, if the clamping distance is interpreted as a reaction coordinate for the conformational exchange process, it is clear that the closing of the mechanical clamp is accompanied by increased curvature in the arms until a maximum twist is reached, after which the arms collapse inward and lose all relative curvature. The maximally twisted state of the arms has precisely the right curvature to wrap around a helical substrate and stabilize kinks in the DNA, suggesting that the propensity to bind DNA is programmed into the intrinsic motions of HU.

dx.doi.org/10.1021/ja4105396 | J. Am. Chem. Soc. 2014, 136, 2204−2207

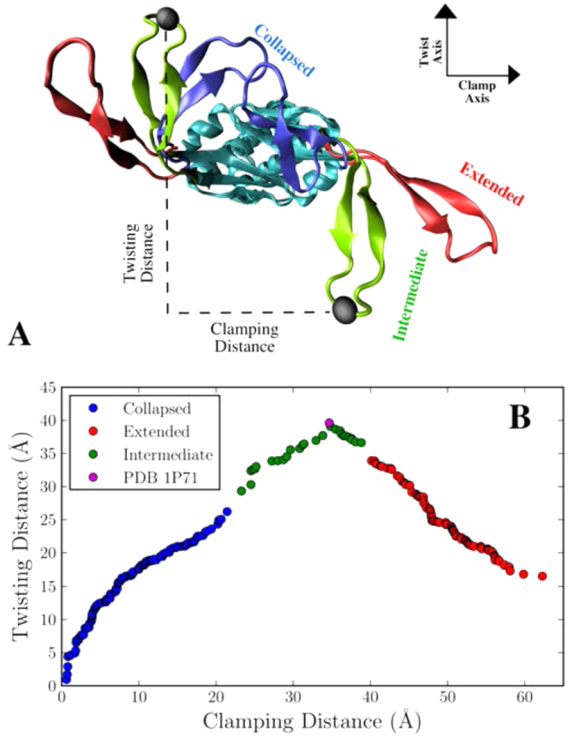

Figure 4. Correlation between the loop motions in HU. (A) Projection of representative structures from the three conformational clusters onto a plane perpendicular to the central axis of the protein. Twisting and clamping distances are defined by the relative position within this plane of the Cα atoms of GLN 64 in the two chains. (B) Twisting distance plotted over the course of the transition from extended to collapsed conformations. The arm curvature reaches a maximum in the intermediate regime corresponding to the bound-like state.

In conclusion, we have described an approach to characterize the conformational fluctuations of proteins based on the use of methyl chemical shifts as structural restraints in replica-averaged MD simulations. By applying this technique to four representative proteins, we hae demonstrated that the approach can be used to generate structural ensembles consistent with NMR measurements. A detailed analysis of the free energy landscape of the HU dimer has enabled us to propose a mechanical clamping mechanism by which this dimeric protein complex can bind efficiently to DNA by transiently sampling in its free state a metastable intermediate with a bound-like structure. This work indicates that Me chemical shifts can serve as particularly sensitive probes of protein dynamics that complement the structural information contained in backbone resonance assignments. Our methodology is generalizable to other proteins and should be broadly useful for understanding the structure and dynamics of biomolecules for which Me chemical shifts have been assigned.

■ ASSOCIATED CONTENT

Supporting Information

Methods, supporting text, Tables S1−S3, and Figures S1−S13. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

mv245@cam.ac.uk

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge a Class 2a grant from the HECToR Supercomputing Service in the UK. A.K. acknowledges a Scholarship from the Winston Churchill Foundation of the United States. C.C. was supported by a Marie Curie IEF fellowship.

■ REFERENCES

(1) (a) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615. (b) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G. H.; Eletsky, A.; Wu, Y. B.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685.

(2) (a) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. *J. Biomol. NMR* **2009**, *44*, 213. (b) Camilloni, C.; Simone, A. D.; Vranken, W. F.; Vendruscolo, M. *Biochemistry* **2012**, *51*, 2224.

(3) (a) Markwick, P. R.; Cervantes, C. F.; Abel, B. L.; Komives, E. A.; Blackledge, M.; McCammon, J. A. *J. Am. Chem. Soc.* **2010**, *132*, 1220. (b) De Gortari, I.; Portella, G.; Salvatella, X.; Bajaj, V. S.; van der Wel, P. C. A.; Yates, J. R.; Segall, M. D.; Pickard, C. J.; Payne, M. C.; Vendruscolo, M. *J. Am. Chem. Soc.* **2010**, *132*, 5993. (c) Lehtivarjo, J.; Tuppurainen, K.; Hassinen, T.; Laatikainen, R.; Peräkylä, M. *J. Biomol. NMR* **2012**, *52*, 257. (d) Robustelli, P.; Stafford, K. A.; Arthur G. Palmer, I. *J. Am. Chem. Soc.* **2012**, *134*, 6365.

(4) Camilloni, C.; Robustelli, P.; Simone, A. D.; Cavalli, A.; Vendruscolo, M. *J. Am. Chem. Soc.* **2012**, *134*, 3968.

(5) Vis, H.; Vorgias, C. E.; Wilson, K. S.; Kaptein, R.; Boelens, R. *Biopolymers* **1996**, *40*, 553.

(6) Vis, H.; Vorgias, C. E.; Wilson, K. S.; Kaptein, R.; Boelens, R. *J. Biomol. NMR* **1998**, *11*, 265.

(7) Robustelli, P.; Kohlhoff, K.; Cavalli, A.; Vendruscolo, M. *Structure* **2010**, *18*, 923.

(8) (a) Cavalli, A.; Camilloni, C.; Vendruscolo, M. *J. Chem. Phys.* **2013**, *138*, 94112. (b) Roux, B.; Weare, J. *J. Chem. Phys.* **2013**, *138*, 084107.

(9) Camilloni, C.; Cavalli, A.; Vendruscolo, M. *J. Phys. Chem. B* **2013**, *117*, 1838.

(10) (a) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923. (b) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 476. (c) Huang, J. R.; Grzesiek, S. *J. Am. Chem. Soc.* **2010**, *132*, 694. (d) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128.

(11) Pitera, J. W.; Chodera, J. D. *J. Chem. Theor. Comput.* **2012**, *8*, 3445.

(12) Sahakyan, A. B.; Vranken, W. F.; Cavalli, A.; Vendruscolo, M. *J. Biomol. NMR* **2011**, *50*, 331.

(13) (a) Schwarten, M.; Stoldt, M.; Mohrüder, J.; Willbold, D. *Biochem. Biophys. Res. Commun.* **2010**, *395*, 426. (b) Klages, J.; Kotzsch, A.; Coles, M.; Sebald, W.; Nickel, J.; Müller, T.; Kessler, H. *Biochemistry* **2008**, *47*, 11930. (c) Vis, H.; Boelens, R.; Mariani, M.; Stroop, R.; Vorgias, C. E.; Wilson, K. S.; Kaptein, R. *Biochemistry* **1994**, *33*, 14858. (d) Soundararajan, M.; Willard, F. S.; Kimple, A. J.; Turnbull, A. P.; Ball, L. J.; Schoch, G. A.; Gileadi, C.; Fedorov, O. Y.; Dowler, E. F.; Higman, V. A.; Hutsell, S. Q.; Sundström, M.; Doyle, D. A.; Siderovski, D. P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6457.

(14) (a) Noda, N. N.; Kumeta, H.; Nakatogawa, H.; Satoo, K.; Adachi, W.; Ishii, J.; Fujioka, Y.; Ohsumi, Y.; Inagaki, F. *Genes Cells* **2008**, *13*, 1211. (b) Stangler, T.; Mayr, L. M.; Willbold, D. *J. Biol. Chem.* **2002**, *277*, 13363. (c) Serban, D.; Arcineigas, S. F.; Vorgias, C. E.; Thomas, G. J. *Protein Sci.* **2003**, *12*, 861.

(15) Swinger, K. K.; Lemberg, K. M.; Zhang, Y.; Rice, P. A. *EMBO J.* **2003**, *22*, 3749.