BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

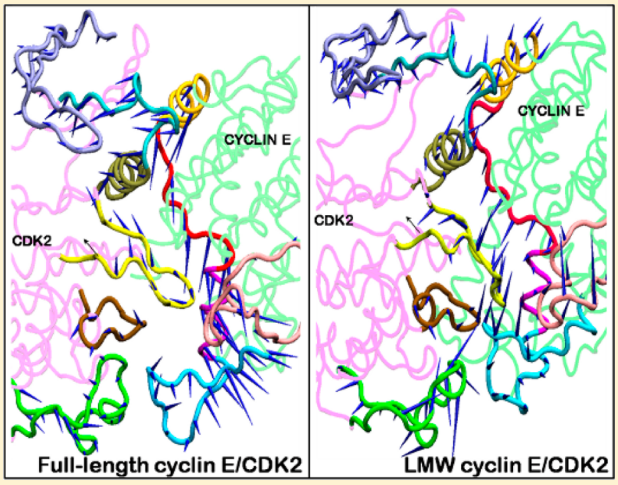
# Why Are the Truncated Cyclin Es More Effective CDK2 Activators than the Full-Length Isoforms?

Soumya Lipsa Rath and Sanjib Senapati*

Bhupat and Jyoti Mehta School of Biosciences, Department of Biotechnology, Indian Institute of Technology Madras, Chennai 600036, India

Supporting Information

**ABSTRACT:** Cell cycle regulating enzymes, CDKs, become activated upon association with their regulatory proteins, cyclins. The G1 cyclin, cyclin E, is overexpressed and present in low molecular weight (LMW) isoforms in breast cancer cells and tumor tissues. *In vivo* and *in vitro* studies have shown that these LMW isoforms of cyclin E hyperactivate CDK2 and accelerate the G1-S phase of cell division. The molecular basis of CDK2 hyperactivation due to LMW cyclin E isoforms in cancer cells is, however, unknown. Here, we employ a computational approach, combining homology modeling, bioinformatics analyses, molecular dynamics (MD) simulations, and principal component analyses to unravel the key structural features of CDK2-bound full-length and LMW isoforms of cyclin E1 and correlate those features to their differential activity. Results suggest that the missing N- and C-terminal regions of the cyclin E LMW isoforms constitute the Nuclear Localization Sequence (NLS) and PEST domains and are intrinsically disordered. These regions, when present in the full-length cyclin E/CDK2 complex, weaken the cyclin-CDK interface packing due to the loss of a large number of key interface interactions. Such weakening is manifested in the decreased contact area and increased solvent accessibility at the interface and also by the absence of concerted motions between the two partner proteins in the full-length complex. More effective packing and interactions between CDK2 and LMW cyclin E isoforms, however, produce more efficient protein−protein complexes that accelerate the cell division processes in cancer cells, where these cyclin E isoforms are overexpressed.




Cyclin dependent kinase 2 (CDK2) is a member of the eukaryotic serine/threonine protein kinase family that assists the phosphoryl transfer of ATP $\gamma$-phosphate to peptide substrates belonging to downstream targets.[1,2] CDK2 participates in cell cycle regulation at the G1/S boundary.[1,3] Its deregulation has been reported to associate with several human tumors, evoking strong interest to understand the molecular mechanism of CDK2 activation.[3,4] The recent structural and mutational studies have suggested that the full activation of CDK2 requires both binding of cyclin A/cyclin E and phosphorylation of residue T160 in the activation segment.[5−7] In the absence of these features, it has negligible activity because the residues that form the catalytic and substrate-recognition sites are disordered.[5,6]

CDK2 contains 298 residues with a small N-terminal domain (residues 1−82), a large C-terminal domain (residues 88−298), and a short hinge region (residues 83−87) connecting the N- and C-terminal domains.[8] The N-terminal domain is primarily composed of a $\beta$-sheet, containing five antiparallel $\beta$-strands, and one $\alpha$-helix. This helix with the "PSTAIRE" sequence (residues 45−56) is a signature of this class of proteins and constitutes the main point of interaction with cyclins.[8,9] The loop which precedes the PSTAIRE helix, known as the 40s loop (residues 33−41), also interacts with cyclins. The C-terminal domain is predominantly $\alpha$-helical and contains the activation segment, the so-called T-loop (residues 145−172). The specific residue Thr160 in the T-loop undergoes phosphorylation by cyclin activating kinase (CAK) for complete CDK2 activation.[7−9] The catalytic cleft that binds ATP is located between the N- and C-terminal domains.[6−9] A glycine rich loop, commonly known as G-loop, lies above the ATP binding pocket and is conserved in many kinases.[8] The primary function of this loop is to align the substrate and ATP correctly for a smooth transfer of the phosphate.[8,9] As a member of the CMGC group of kinases, CDK2 has a CMGC-specific structural feature, such as the CMGC insert at the CDK2-cyclin interface.[2]

Cyclin E is the regulatory subunit of CDK2 in the late G1 phase.[10] The isoform, cyclin E1, contains 395 amino acid residues and is expressed in most proliferating normal and tumor cells.[10,11] The structure of cyclin E1 is primarily $\alpha$-helical

**Received:** April 4, 2014
**Revised:** June 20, 2014
**Published:** June 20, 2014

   dx.doi.org/10.1021/bi5004052 | *Biochemistry* 2014, 53, 4612−4624

and comprises two five-helical domains with additional helices at the N- and C-termini. The N-terminal domain consists of the N-terminal-helix ($\alpha$Nter) followed by five helices, $\alpha1-\alpha5$. The hydrophobic core of this cyclin box fold is formed by helix $\alpha3$, which is surrounded by the other four helices. Similarly, the C-terminal domain comprises five helices $\alpha1'-\alpha5'$ and the C-terminal-helix ($\alpha$Cter).[12] The protein also possesses a nuclear localization sequence (NLS) at the N-terminal end and a PEST sequence (a sequence that is rich in proline (P), glutamic acid (E), serine (S), and threonine (T)) at the C-terminal end for targeting the protein for ubiquitination.[10,12−15]

Cyclin E (both isoforms E1 and E2) is reported to have a profound role in oncogenesis.[10,16] It is found to be overexpressed and present in low molecular weight (LMW) isoforms in breast cancer cells and tumor tissues.[17−23] In normal cells, however, cyclin E is expressed precisely when needed and present in full-length form.[18,20] The mass of these LMW isoforms ranges between 49 to 33 kDa, and they are typically excluded from 25−100 residues from the N-terminal end.[17,18,21] Porter et al. have shown that the generation of the tumor specific LMW forms of cyclin E is predominantly derived from proteolytic processing of the full-length cyclin E.[22,23] *In vitro* and *in vivo* studies have shown that the LMW forms of cyclin E are not only functional but also more active (in activating CDK2) than the full-length cyclin E.[12,17,22,23]

The crystal structure of the cyclin E/CDK2 complex has been solved only very recently.[12] The structure contains a truncated version of cyclin E1 with residues 81−363. Thus, it lacks the two important recognition sites: the NLS and the PEST sequence. However, both the N- and C-terminal cyclin boxes are present in the construct, and the protein associates tightly with CDK2. The structural elements of CDK2 are found to be in active conformation, and the overall structure resembles the active cyclin A/pCDK2 complex. The kinetic analyses show that this truncated complex is a more effective enzyme for the phosphorylation of peptide substrates than the full-length cyclin E1/CDK2 complex.[12] Even though the structure of a truncated cyclin E/pCDK2 complex has been obtained recently, the literature still lacks the structure of a full-length cyclin E/pCDK2 complex. More importantly, the structural basis of higher activity of the truncated cyclin E/CDK2 complexes over the full-length cyclin E/CDK2 complex is yet to be understood. Here, we employ a computational approach, combining homology modeling, molecular dynamics (MD) simulations, bioinformatics, and principal component analyses to unravel the key structural features of CDK2-bound full-length cyclin E1 and correlate those features to the differential activity of truncated and full-length cyclin E1/CDK2 complexes.

## MATERIALS AND METHODS

The structure of full-length cyclin E1 (residues 1−395)/CDK2 was modeled from the crystal structure of the truncated cyclin E1 (residues 81−363)/CDK2 complex. Thus, the atomic coordinates of residues 81−363 were obtained directly from the Protein Data Bank (PDB ID: 1W98; resolution 2.15 Å).[12] The missing sequences of residues 1−80 and 364−395 were obtained from Universal Protein Resource database (UniProt entry: P24864-3).[24] *Ab initio* protein structure predictor, QUARK was used to construct 3D models of these sequences, after noting a very low sequence similarity with available structures in databases (~23%).[25,26] QUARK relies on available experimental structures of small fragments of 20 residues, which it assembles by replica exchange Monte Carlo simulations using an optimized knowledge-based force field. These modeled structures of N- and C-termini were merged with the crystal conformation of the truncated cyclin E1/CDK2 complex and the resultant structure was optimized by the conjugate gradient energy minimization method in Modeler program.[27] Before a more exhaustive refinement of this model by MD simulations, the stereochemical quality (i.e., $\varphi$, $\psi$ distribution) and physical correctness (e.g., number of correct bonds/angels) were assessed by using Procheck[28] and MolProbity[29] programs, which perform structural analyses based on the available protein crystal structures in databases (Table S1, Supporting Information). The structure of the second truncated complex, cyclin E1 (residues 101−395)/CDK2 was generated from the full-length complex by deleting the N-terminal 100 residues. An ATP and a $Mn^{2+}$ ion were included in each model structure based on their locations in the cyclin A/CDK2 crystal structure; PDB ID, 1JST.[7]

In each of the above structures, hydrogens for heavy atoms were added by the leap program in AMBER 12.0.[30] Subsequently, energy minimization for 1000 steps using the conjugate gradient and another 1000 steps using the steepest descent algorithm were carried out. The protonation states of histidines, HID or HIE, were determined by the local hydrogen bonding network using the WHATIF program.[31] After relaxing the added atoms in gas phase, the structures were solvated in cubic periodic boxes of explicit water with water molecules extending 9 Å outside the protein−protein complexes on all sides. The 3-site TIP3P model was chosen to describe the water molecules.[32] To neutralize the systems, seven, nine, and six $Na^+$ ions were added to the Trunk 1, full-length, and Trunk 2 complexes, respectively. The parameters for ATP, $Mn^{2+}$, and phosphorylated threonine (T160 in CDK2) were obtained from an AMBER contributed parameter data set.[33−35] An extensive set of minimization and thermalization was performed on each complex by maintaining harmonic restraints on heavy atoms of the crystallized domains, followed by gradual heating to 300 K in canonical ensembles. The harmonic restraints were gradually reduced to zero, and solvent density was adjusted under isobaric and isothermal conditions at 1 atm and 300 K. The systems were equilibrated for 5 ns in the NPT ensemble, with a simulation time step of 2 fs. During this period, the RMSDs were seen to converge in about 40 ns for full-length and 15 ns for the truncated complexes. The full-length structure was simulated for 200 ns, and the truncated complexes were simulated for 100 ns each. Besides, two independent simulations on the QUARK generated models of N- and C-terminal domains were carried out for microseconds, following the same protocol. Table 1 lists all the systems that are studied in this work. The long-range electrostatic interactions were treated by using the particle-mesh Ewald sum, and SHAKE was

**Table 1. List of Systems Studied**

| serial no. | system | duration of simulation |
|---|---|---|
| 1 | pCDK2/cyclin E1(1−395) | 200 ns |
| 2 | pCDK2/cyclin E1(81−363) | 100 ns |
| 3 | pCDK2/cyclin E1(101−395) | 100 ns |
| 4 | cyclin E1(1−80) | 5500 ns |
| 5 | cyclin E1(364−395) | 1500 ns |
| 6 | pCDK2/cyclin E1(81−363) (modeled from full-length) | 50 ns |


dx.doi.org/10.1021/bi5004052 | *Biochemistry* 2014, 53, 4612−4624

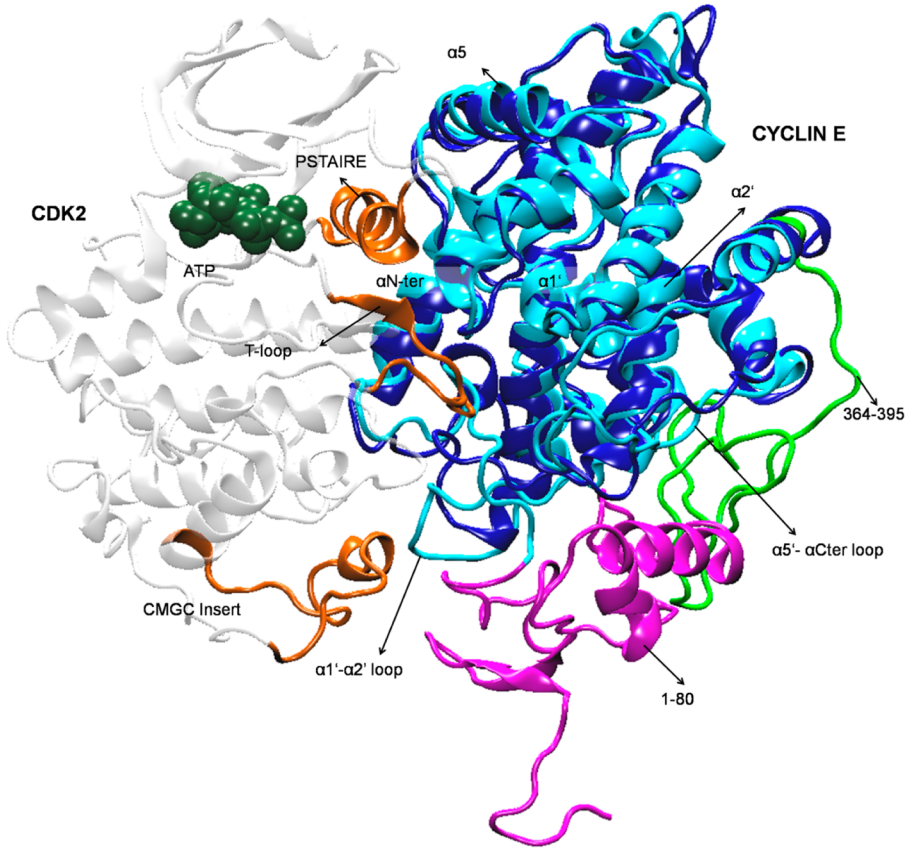



**Figure 1.** Structural comparison between the full-length and truncated cyclin E1/CDK2 complexes. Time-averaged structure of the model full-length cyclin E1/CDK2 complex is superposed on the crystal structure of a truncated complex (crystal structure PDB ID: 1W98, resolution 2.15 Å). Color scheme: crystal cyclin E1 (81−363), dark blue; model full-length cyclin E1 (1−80), magenta; (81−363), cyan; and (364−395), green. CDK2 in the full-length complex is shown in gray and is omitted from the crystal structure as it was similar. Important interface regions of CDK2 are highlighted in orange. ATP is shown in van der Waals mode in dark green.

used to constrain bond lengths between heavy atoms and hydrogens.[36,37] All simulations were performed using the AMBER 12.0 molecular dynamics simulation program with the AMBER ff99SB force field.[38]

Principal component analysis (PCA) is utilized here to study the large concerted motions. PCA transforms the original high-dimensional representation of protein motions into a low-dimensional one that captures the dominant modes of the protein motions.[39,40] The analysis involves two main steps: (i) calculation of the covariance matrix C of the positional deviations and (ii) diagonalization of this matrix. The 3N dimensional covariance is calculated based on an ensemble of protein structures, whose elements are defined as

$$C_{ij} = \left\langle (x_i - \langle x_i \rangle) \cdot (x_j - \left\langle x_j \right\rangle) \right\rangle$$

where, $x_i$ and $x_j$ are the atomic coordinates, and the brackets denote ensemble average. The diagonalization of the symmetric matrix $C$ is equivalent to solving the eigenvalue problem:

$$A^T CA = \lambda$$

where A represents the eigenvectors, and $\lambda$ represents the associated eigenvalues. In a sense, the eigenvectors are directions in configurational space that represent collective motions. Corresponding eigen values define the mean square fluctuation of the motion along these vectors. The collective variables are ranked according to their contribution to the total mean-square fluctuation. Prior to the calculation of the covariance matrix, we removed the overall rotation and translation. PCA analysis was carried out using the PCAZIP suite of programs.[41]

## RESULTS AND DISCUSSION

**Model Structure of the Full-Length Cyclin E1/CDK2 Complex.** In MD simulations, the added domains of the full-length complex are expected to attain the native, folded conformations by virtue of their intrinsic dynamics, as described by intra- and intermolecular interactions in solution.[42,43] However, it is always necessary to ensure that the system is well equilibrated before performing the analyses on simulation trajectory. A simple strategy to ensure this is to monitor the mean squared displacements of the protein residues from their initial positions.[39,44] Supporting Information, Figure S1 presents the root mean squared displacements (RMSD) of the full-length cyclin E1/CDK2 complex for its C$\alpha$ atoms. The figure shows that the RMSDs converge only at 40 ns, indicating that the system required a very long simulation time to equilibrate. Such a long equilibration time is not unexpected considering the size of the added N- and C-terminal regions and also the overall large size of the protein−protein complex. The subsequent analyses of the full-length complex were therefore performed on the final 160 ns simulation data.

Thus, the time-averaged structure of the full-length complex was generated by averaging over the 160 frames separated by 1 ns from the postequilibration phase. Figure 1 shows the time-


dx.doi.org/10.1021/bi5004052 | Biochemistry 2014, 53, 4612−4624

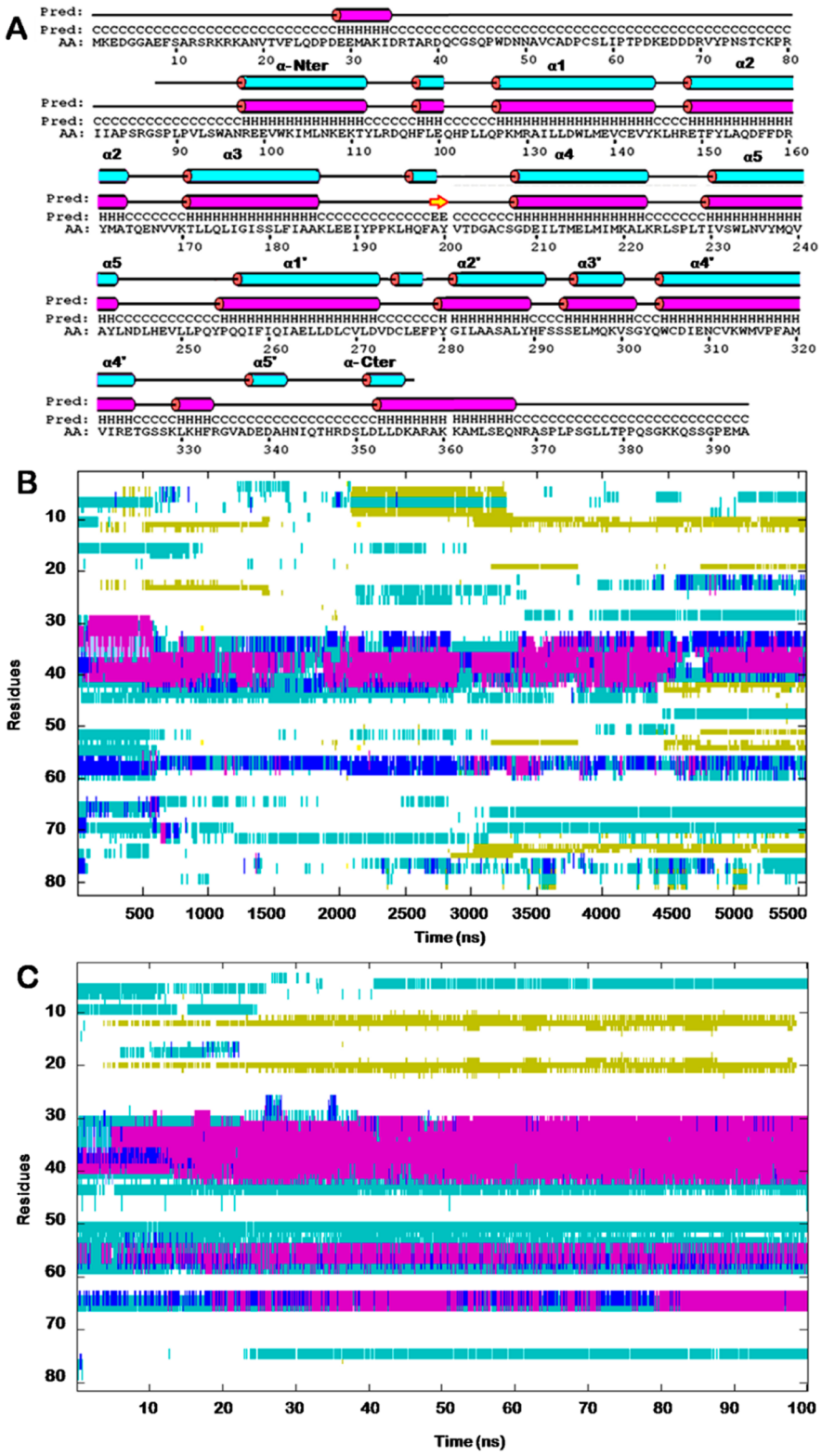



**Figure 2.** Missing domains in cyclin E are intrinsically disordered. (A) Predicted secondary structural elements of full-length cyclin E (pink) are compared with that of the crystal structure of truncated cyclin E (cyan). A *de novo* structure predictor could correctly reproduce the secondary elements of the cyclin E central region (residues 81−363) present in the crystal structure. (B) Time-evolution of the secondary structures of the added N-terminal region, residues 1−80, during 5.5 $\mu$s of independent simulation. (C) Time-evolution of the same region, residues 1−80, but when bound to the central cyclin E region. In bound form, an increasing tendency to form new secondary elements is evident. Color scheme: $\alpha$-helix, purple; 3−10 helix, blue; $\beta$-sheet, yellow; turn, cyan; and coil, white.

averaged structure of the model full-length cyclin E1/CDK2 complex and its comparison with the crystal structure of the truncated complex. The figure shows that the central cyclin domain (residues 81−363), including the helices $\alpha$Nter, $\alpha$3, $\alpha$4,


dx.doi.org/10.1021/bi5004052 | *Biochemistry* 2014, 53, 4612−4624

and $\alpha$5 that are responsible for making interactions with CDK2, remains largely unaffected in the full-length complex. The time-averaged RMSD of the central domain was found to be only 1.78 Å. However, certain other primary contact regions, e.g., loops $\alpha$1′-$\alpha$2′ and $\alpha$5-$\alpha$1′, and helix $\alpha$1′, show significant deviations from the crystal structure of the truncated complex. The time-averaged RMSD of the respective region was noted to be 4.5 Å, 2.2 Å, and 3.4 Å. These regions were found to be shifting away from the interface, weakening cyclin-CDK2 interactions in the full-length complex. A detailed comparison of the cyclin-CDK interface, in these complexes, is presented in the next section.

**Missing Domains in Cyclin E1 Are Intrinsically Disordered.** The weakening of the interfacial interactions in the full-length complex implies a possible role of the added N- and C-terminal regions on the stability of this complex. Hence, a detailed investigation was performed on these added regions. The relative internal motions of the protein residues, which can be described from the isotropic temperature factor (*B*-factor), show significantly higher flexibility of the added regions than the central 81−363 region. This is shown in Figure S2 (Supporting Information). The figure also shows that the simulation results could reproduce the crystal structure *B*-factor values for the central region.[12] For example, the loops $\alpha$1′-$\alpha$2′ and $\alpha$5-$\alpha$1′, and helix $\alpha$3, etc. show high fluctuations in both cases.

Structure prediction and independent simulations performed on the added regions show no or minimal secondary structures (Figure 2A,B). However, while bound to the central region in the full-length complex, these domains, particularly the N-terminal domain shows an increased tendency to form secondary structures (Figure 2C). As shown in Figures 1 and 2C, the N-domain forms a long and stable $\alpha$-helix, two small helices, and an antiparallel $\beta$ sheet in bound form. The high flexibility, low secondary structure content in the unbound form, and induced folding upon binding to a partner are some of the general structural characteristics of intrinsically disordered proteins (IDPs) or of proteins with intrinsically disordered regions (IDRs).[45,46] This led us to examine these structures for IDRs.

Bioinformatics analyses, which can predict disordered regions in proteins with accuracy levels higher than 70% are gaining tremendous importance.[47−49] This is particularly because (i) nearly 75% of the signaling proteins and 25% of all proteins in eukaryotes are predicted to contain intrinsically disordered regions.[47] Moreover, among the solved structures in the PDB, about 70% contain some disordered residues, while 25% contain IDRs with >10 residues. (ii) Biophysical studies of IDPs or IDRs possess tremendous experimental challenges due to their broad conformational heterogeneity.[48,49] Here, we adopt recent computational approaches to show that the added regions are highly unstructured. Apart from the above-mentioned characteristics, we identified several other features, typical of inherently disordered proteins (IDPs), for both regions. These were low hydrophobicity, a biased amino acid composition with enrichment in the disorder promoting residues, a little or no well-defined secondary structure, etc.

Analysis of the hydropathic character of the proteins by the Kyte-Doolittle method, which derives hydrophobicity of the amino acids from experimental data of partitioning of peptides in apolar and polar solvents, indicates that both added regions are highly hydrophilic in nature (Figure S3, Supporting Information).[50] The amino acid sequence itself, also indicates that both domains are rich in hydrophilic residues (Table 2). For example, the N-terminal region (residues 1−80) contains

**Table 2. Distribution of Hydrophilic Residues along the Cyclin E1 Sequence**

| | 1−80 | 81−363 | 364−395 |
|---|---|---|---|
| Ser + Thr | 11.25% | 9.00% | 17.9% |
| Lys + Arg | 16.25% | 8.60% | 17.9% |
| Glu + Asp | 20.00% | 9.40% | 5.10% |
| Gly | 3.75% | 2.80% | 7.60% |
| total | 51.25% | 29.8% | 48.5% |

20% Glu+ Asp, 16% Lys + Arg, and 11% Ser + Thr. It also contains 4% Gly, which therefore would favor a solvent-exposed, flexible unfolded structure. In contrast, the central region (residues 81−363) has less hydrophilicity and contains a total of only 29.8% of such residues.

Prediction of protein disorder by a consensus artificial neural network prediction method, PONDR-FIT, indicates that majority of the residues in added regions have higher disorder probability than the central region (Figure S4, Supporting Information).[51] These disordered regions possess a limited amount of secondary structure as indicated by the structure predictor, PSIPRED, which is also a neural network based prediction method.[52] This is shown in Figure 2A. Figure 2A also included a comparison of the secondary structural elements as predicted by PSIPRED and as present in the crystal structure of cyclin E1. A very good match of the secondary structures of the central region validates the *in silico* disorder predictions. Finally, following the more recent studies that have established that a combination of low mean hydrophobicity and relatively large net charge represents a unique structural feature of IDPs/IDRs and that these proteins are localized within a certain region of charge-hydrophobicity (C−H) phase space,[45,46,53] we found the following. The added N- (blue) and C-terminal (green) domains of full-length cyclin E1 are localized in the disordered region of the C−H plot, while the central region (red) is well placed in the ordered region (Figure 3). Our finding is consistent with the fact that (i) the NLS and PEST domains of most of the proteins are reported to be intrinsically disordered and that (ii) the terminal regions of cyclin E could not be captured in the electron density map, presumably due to their random motions. The classification boundary between disordered and ordered proteins in the C−H plot was determined from a large training set of proteins.[45,46,53,54]

At this point, we are tempted to comment on the nature of binding of the disordered terminal regions to the truncated CDK2/cyclin E complex, from the viewpoint of existing mechanistic models. Among the proposed mechanisms of binding of IDPs to their biological targets, "conformational selection" and "coupled folding and binding" have received much acceptance.[55−57] While the first mechanism assumes that the folding of IDPs would precede binding, the second mechanism argues conformational transitions to folded forms only after binding to the targets. Our results of cyclin E terminal domains attaining secondary structures only after binding to the CDK/cyclin complex (Figure 2) augurs very well with the "coupled folding and binding" mechanism of IDPs.

**Intrinsically Disordered Cyclin E1 Regions Weaken the Cyclin E1/CDK2 Interface.** After establishing that the missing terminal regions of cyclin E1 are IDRs, we proceeded


dx.doi.org/10.1021/bi5004052 | *Biochemistry* 2014, 53, 4612−4624

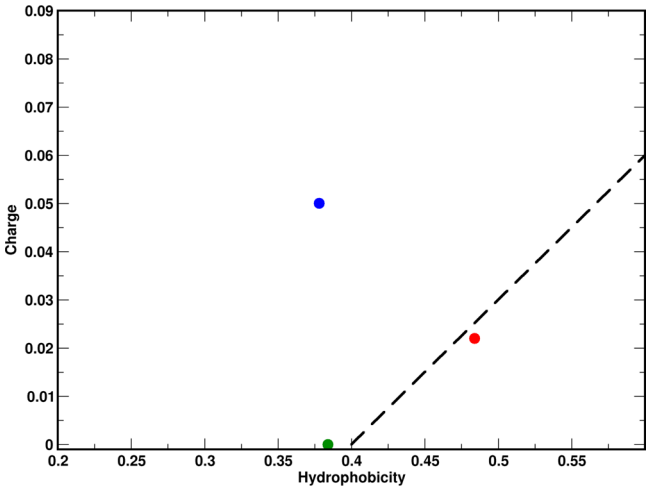



**Figure 3.** Comparison of the mean net charge versus mean hydrophobicity of different cyclin E domains. Results are shown for residues 1−80 (blue), 81−363 (red), and 364−395 (green). The dashed line represents the boundary between disordered (right) and ordered proteins (left), obtained from a large data set.

to check their effects on the cyclin E1-CDK2 interface. The findings from the full-length complex are compared with the simulation results from control systems that comprise truncated cyclin E/CDK2 complexes. As the first control, we simulated the crystal structure of the truncated complex that has cyclin E1 with residues 81−363 (Trunk 1). The second control system comprises CDK2, complexed with a physiologically more relevant variant of cyclin E1 with residues 101−395 (Trunk 2). This 33.9 kDa isoform is reported to be overexpressed in tumor cells and tissues and activates CDK2 more effectively than full-length cyclin E.[17,18] We simulated both the control systems for 100 ns each. The details of the systems are presented in Table 1.

The RMSD comparison of the complexes implies that the full-length complex undergoes larger conformational changes than the truncated complexes, Trunk 1 and Trunk 2 (Figure S5, Supporting Information). This is not surprising recalling the fact that cyclin E in the full-length complex contains two IDRs. Both the truncated complexes stabilize relatively quickly, with Trunk 1 reaching a plateau within 15 ns of simulation time. The quick equilibration of Trunk 1 is not unexpected as this system represents the crystal structure. The RMSD of Trunk 2 is very similar to Trunk 1, except for the initial phase during which the additional C-terminal domain in Trunk 2 adjusts to the environment of the cyclin E crystal conformation. Since the C-terminal domain is not very close to the cyclin E/CDK2 interface, its inclusion appears to have less influence on the overall stability of Trunk 2. The *B*-factor analysis for the common cyclin E region in all three complexes (residues 101−363), also implies higher fluctuations of residues in the full-length complex (Figure S6, Supporting Information). Loop $\alpha1'$-$\alpha2'$ (residues 244−254) that comprises some of the primary contact residues for the interactions with CDK2 shows significantly higher fluctuations in full-length cyclin E. Another principal contact, $\alpha5$ that packs parallel to PSTAIRE helix of CDK2 in truncated complexes shifts out to some extent in the full-length complex due to the loss of some important contacts (details below). As a result, the adjacent $\alpha5$-$\alpha1'$ loop also becomes weakly packed in this complex. The N-terminal helix and $\alpha$Nter-$\alpha1$ loop also show higher fluctuations in the full-length complex than the Trunk 2 complex. The large $\alpha5'$-$\alpha$Cter loop, which protrudes toward the N-terminal in all complexes (as similar to crystal structure), however, show reduced flexibility in the full-length complex due to the favorable interactions with the added N-terminal region. Correspondingly in CDK2, the T-loop, the extended T-loop region, and the PSTAIRE helix region, which interact strongly with cyclin interface residues via van der Waals and electrostatic interactions, were found to have higher *B*-factors in the full-length complex (data not shown).

The cyclin E/CDK2 interface has also been compared from the time-averaged simulated structures of the three complexes. Results for the cyclin E/CDK2 interface contacts in Figure 4 suggest that the full-length cyclin E has lesser contacts with CDK2 than the truncated isoforms. The number of contacts is calculated to be 72, 81, and 91 in the full-length, Trunk 1, and Trunk 2 complexes, respectively. Thus, the interface of truncated complexes has an additional 13% and 26% contacts. To recall, the pCDK2/cyclin E1 interface packing is the key to the activating properties of cyclins to CDKs.[6,7] As Figure 4 indicates, lesser contacts in the full-length complex primarily stem from weaker interactions of helix $\alpha1'$ and loop $\alpha1'$-$\alpha2'$ of cyclin E with the T-loop region of CDK2 and that of the $\alpha5$-$\alpha1'$ loop and C-terminal end of the $\alpha3$ helix of cyclin E with the PSTAIRE helix of CDK2. Helices $\alpha5$ and $\alpha$Nter also show loss of contacts with CDK2 interface residues, in the full-length complex. This is consistent with larger *B*-factor values of these regions in the full-length complex in Figure S6 (Supporting Information). The contact map is obtained by counting the number of residue−residue contact areas that exceed a threshold value of 5 Å$^2$.[58] The interfacial packing is also examined in terms of solvent accessible surface area (SASA) at the cyclin-CDK interface. The SASA is computed to be 102.82 nm$^2$ for the full-length complex, 85.69 nm$^2$ for the Trunk 1 complex, and 73.77 nm$^2$ for the Trunk 2 complex, i.e., an increase of 39% of the full-length from the Trunk 2 complex (Figure 4D). This increased solvent exposure could weaken the protein−protein interactions and lower the activation of CDK2 by cyclin E.

To quantify the weakening of protein−protein interactions in the full-length complex, we calculated the average interaction energy between cyclin E and CDK2 over the final 25 ns simulation trajectory. The interaction energy of both truncated complexes was found to be much greater than that of the full-length complex (Figure 5A). The physiologically relevant Trunk 2 complex was found to be the most stable, with an energy gain of ∼150 kcal/mol compared to that of the full-length complex. To get a deeper understanding on such energy differences, residue-level contribution to total energy is analyzed, and the results are shown in Figure 5B. As expected, the major contribution to interprotein interactions came from the interface residues. Particularly, the residues that belong to the secondary elements, loops $\alpha1'$-$\alpha2'$, $\alpha5$-$\alpha1'$, and $\alpha5'$-$\alpha$Cter, and helices $\alpha1'$, $\alpha5$, $\alpha$Nter, and $\alpha3$, contributed most favorably toward better stabilization of the truncated complexes. This is consistent with the contact map data in Figure 4. The cyclin residues that showed differential binding strength in these complexes include Val250, Leu251, and Pro253 from the $\alpha1'$-$\alpha2'$ loop; Val237 from the $\alpha1'$ helix; Arg225 from $\alpha5$; Val237 from the $\alpha1'$ helix; Arg225 from the $\alpha5$-$\alpha1'$ loop; Glu340, Asp341, and Gln344 from the $\alpha5'$-$\alpha$Cter loop etc. (indicated by an * in Figure 5B). The primary contact residues from CDK2 include Phe152, Val154, Val156, Arg157, and Phe179 from the


dx.doi.org/10.1021/bi5004052 | *Biochemistry* 2014, 53, 4612−4624

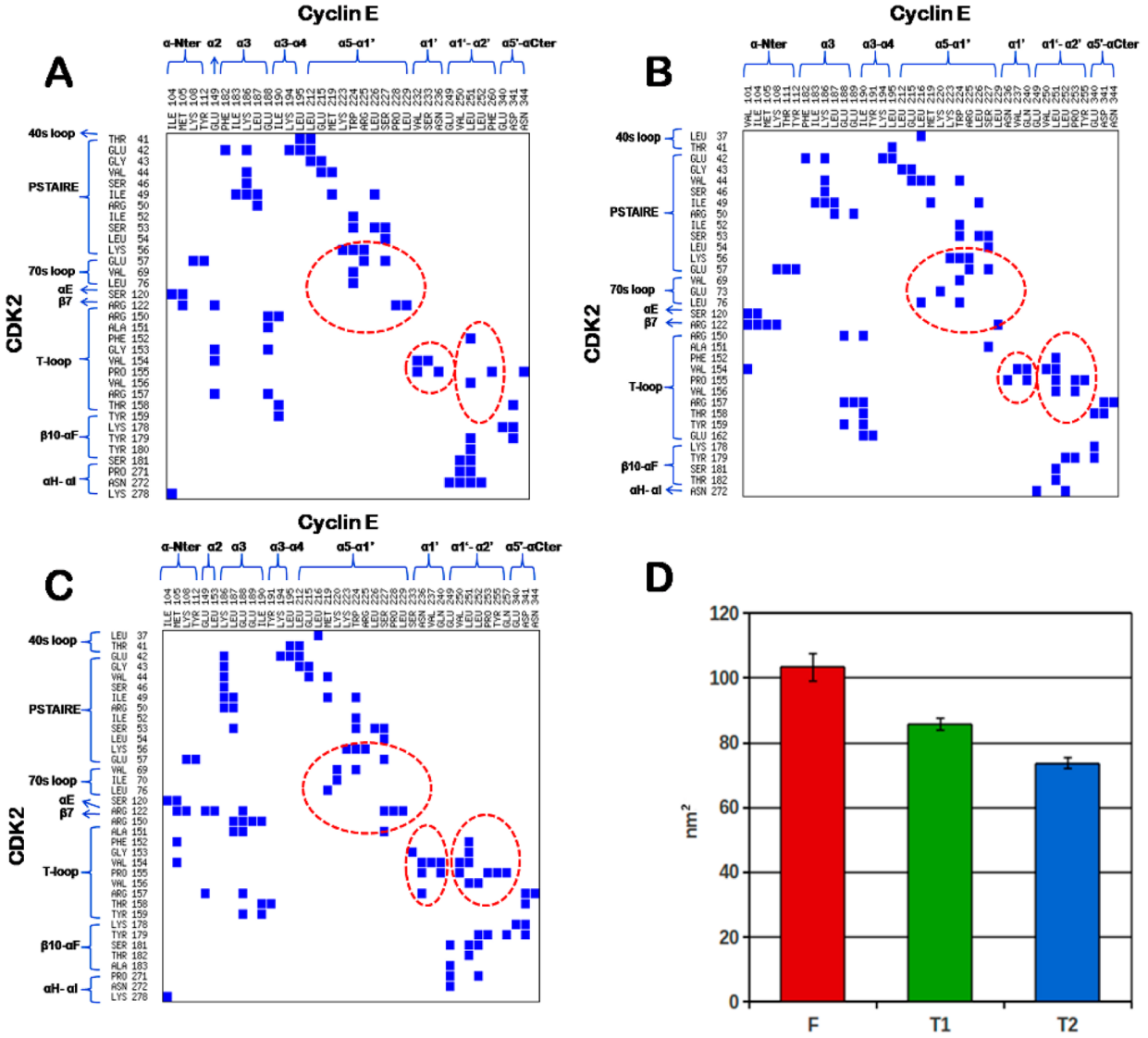



**Figure 4.** Comparison of the cyclin E/CDK2 interface. Contact maps are shown for (A) full-length, (B) Trunk 1, and (C) Trunk 2 complexes. Each square represents contact between a pair of residues. Secondary elements that exhibited significant difference in contacts are highlighted by red circles. (D) Solvent accessible surface area (SASA) at the cyclin E/CDK2 interface of full-length (red), trunk 1 (green), and trunk 2 (blue) complexes.

T-loop and Ser46, Glu57, etc. from the PSTAIRE helix region, etc. (indicated by an * in Figure S7, Supporting Information).

Figure 6 shows the structural changes in the proteins that caused the loss of contacts in the full-length complex. The van der Waals interactions of the side chains of cyclin E:Val250$_{\alpha1'\text{-}\alpha2'}$ and Leu251$_{\alpha1'\text{-}\alpha2'}$ with CDK2:Phe152$_{\text{T-loop}}$ and Val156$_{\text{T-loop}}$ that exist in the crystal structure and constantly persisted in the simulated structures of truncated complexes became significantly weaker in the full-length complex. The hydrogen bond, present between the main chain oxygen of cyclin E:Leu251$_{\alpha1'\text{-}\alpha2'}$ and the hydrogen of main chain nitrogen of CDK2:Val156$_{\text{T-loop}}$ in both truncated complexes, was found to be lost in the full-length complex. These changes are shown in Figure 6A. The stacking of cyclin E:Pro253$_{\alpha1'\text{-}\alpha2'}$ with CDK2:Tyr179$_{\text{T-loop-}\alpha\text{F}}$, as seen in the crystal structure and implicated in both truncated complexes, was also lost in the full-length complex (Figure 6B). The hydrophobic interactions between the side chains of cyclin E:Val237$_{\alpha1'}$ and CDK2:Val154$_{\text{T-loop}}$ became weaker in the full-length complex, due to an increase in residue−residue distance, compared to those in the crystal structure and truncated complexes (not shown). The $\alpha$5-$\alpha$1′ and $\alpha$5′-$\alpha$Cter loop regions also show distinct differences in interaction with CDK2. In the full-length complex, cyclin E:Arg225$_{\alpha5\text{-}\alpha1'}$ shows weaker interactions with CDK2:Glu57$_{\text{PSTAIRE}}$ due to the slight displacement of the $\alpha$5 helix, which packs parallel to the PSTAIRE helix in truncated complexes. The strong hydrogen bonding network involving cyclin E:Asp341$_{\alpha5'\text{-}\alpha\text{Cter}}$, Asn344$_{\alpha5'\text{-}\alpha\text{Cter}}$, and CDK2:Arg157$_{\text{T-loop}}$ side chains in truncated complexes was lost in the full-length complex due to the shifting of the $\alpha$5′-$\alpha$Cter loop toward the added N-terminal region (Figure 6C). Likewise, the electrostatic interactions that were present between cyclin E:Glu340$_{\alpha5'\text{-}\alpha\text{Cter}}$ and CDK2:Lys178$_{\text{T-loop-}\alpha\text{F}}$ in truncated complexes became weaker in the full-length complex due to a larger separation of the residues. The weakening of these interactions in the full-length complex mainly stem from the added N- and C-terminal regions, which due to their intrinsic disorder nature introduce larger dynamics into the adjacent secondary


dx.doi.org/10.1021/bi5004052 | *Biochemistry* 2014, 53, 4612−4624

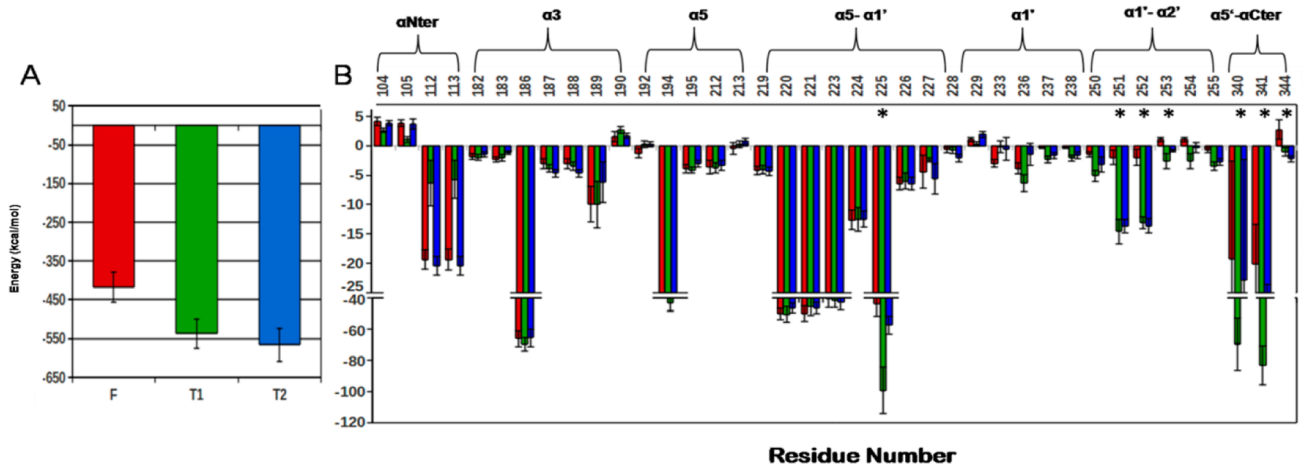



**Figure 5.** Energy of interaction between CDK2 and cyclins. (A) Time-averaged protein−protein interaction energy in full-length (red), Trunk 1 (green), and Trunk 2 (blue) complexes. (B) Contribution of individual cyclin E interface residues toward total energy. Important contributors are highlighted by an *.

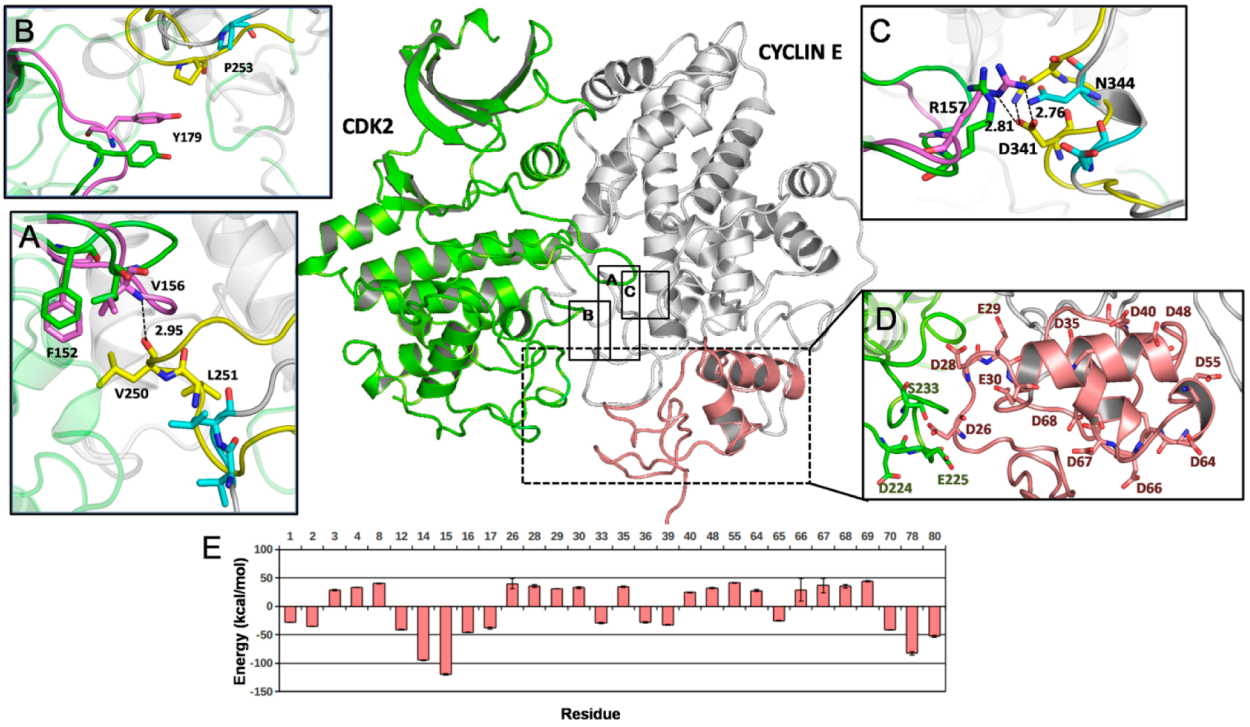



**Figure 6.** Structural changes in the full-length cyclin E/CDK2 interface compared to those of the truncated complex, Trunk 1. (A) The H-bond present between interface residues Leu251$_{\alpha1'\text{-}\alpha2'}$ of cyclin E (yellow) and Val156$_{\text{T-loop}}$ of CDK2 (pink) in Trunk 1 was lost in the full-length complex. Full-length cyclin E and bound CDK2 are shown in cyan and green, respectively. Residues are shown in licorice mode and colored according to atom type. (B) The observed ring stacking between cyclin E:Pro253$_{\alpha1'\text{-}\alpha2'}$ and CDK2:Tyr179$_{\text{T-loop-}\alpha\text{F}}$ in Trunk 1 was seen to be absent in the full-length complex. (C) The hydrogen bonding network in Trunk 1, involving cyclin E:Asp341$_{\alpha5'\text{-}\alpha\text{Cter}}$, Asn344$_{\alpha5'\text{-}\alpha\text{Cter}}$, and CDK2:Arg157$_{\text{T-loop}}$, was also lost in the full-length complex. (D) In the full-length complex, the added N-terminal domain reaches the CDK2:CMGC insert domain (green) for interaction. (E) Residue-level energy contribution of the cyclin E domain, residues 1−80, toward its interaction with the CMGC insert domain.

elements. Moreover, the added N-terminal region can interact with the CMGC insert domain of CDK2 as shown in Figure 6D. This interaction is mostly repulsive due to the richness of charged residues in both the interacting domains (Figure 6E), which also destabilizes the complex to some extent.

The relative strains in folding as the binding of the terminal region to the truncated cyclin E/CDK2 complex occurs were estimated to be 547.03 and 310.65 kcal/mol, respectively, for N- and C-terminal regions. Relative strains were computed from the distribution of molecular mechanics generalized Born surface area (MMGBSA) energies of the uncomplexed form as compared to the terminal regions in the complexed form ($\Delta G_{\text{strain}} = \Delta G_{\text{complex}} - \Delta G_{\text{truncated}} - \Delta G_{\text{terminal}}$). Thus, it is likely that crystal contacts and shorter constructs may lead to the observed contacts, and in the cell, the interaction is weak due to complex kinetics.

**Essential Motions of Full-Length and Truncated Complexes Differ.** We performed principal component analysis (PCA) to locate the large concerted motions, present between cyclin E and CDK2. PCA transforms the original space


dx.doi.org/10.1021/bi5004052 | *Biochemistry* 2014, 53, 4612−4624

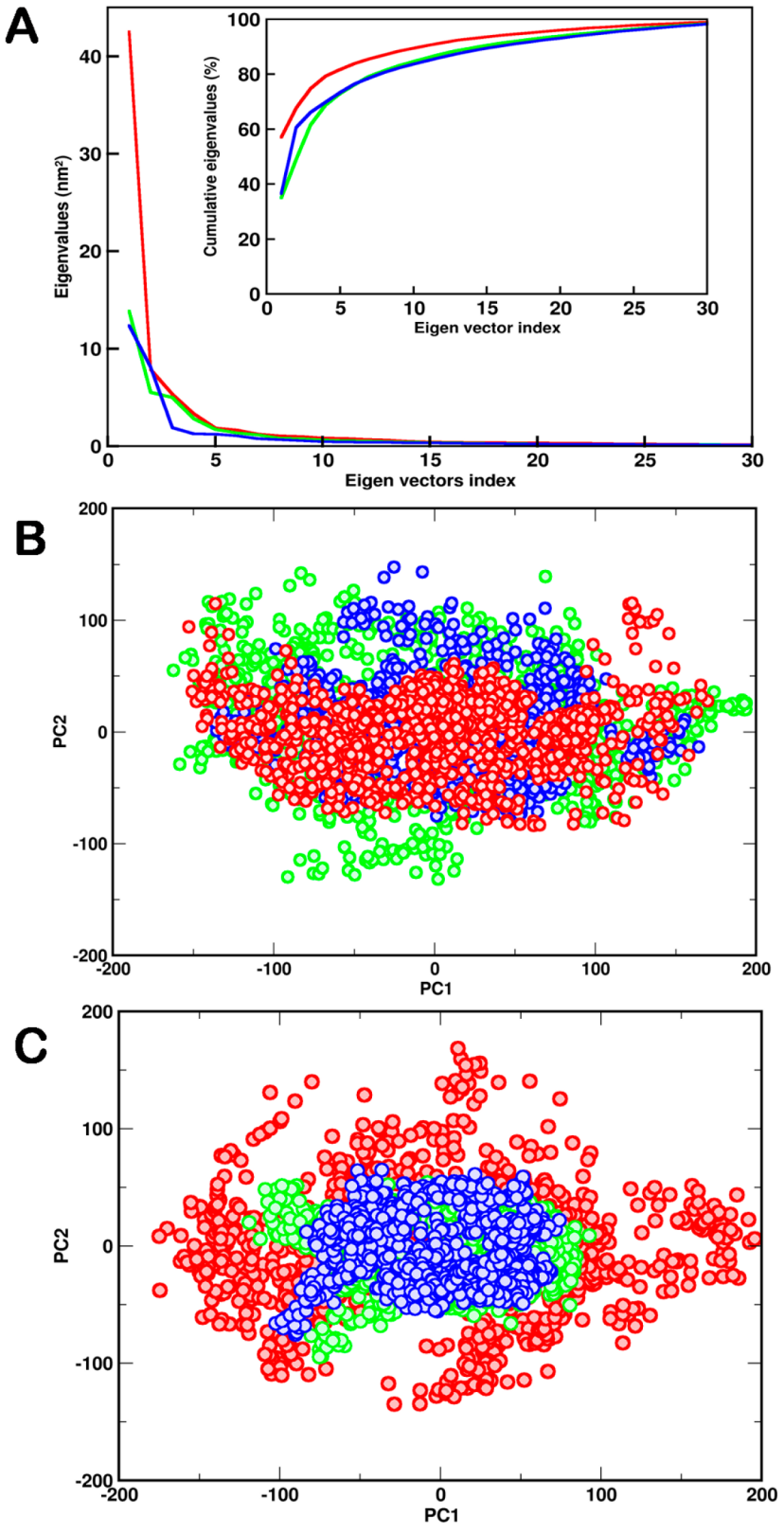



**Figure 7.** Eigenvalue profile and comparative sampling of essential motions in three cyclin E/CDK2 complexes. (A) Eigenvalues of the first 30 eigenvectors, derived from PCA analyses of the simulation trajectories of the full-length (red), Trunk 1 (green), and Trunk 2 (blue) complexes, are shown. The inset shows the cumulative contribution of these eigenvectors. (B and C) Two dimensional projection of the simulated structures on the plane constituted by the first two principal components, in all three complexes: (B) for CDK2 and (C) for cyclin E.

of correlated variables into a reduced space of independent variables (the principal components) and thus helps in locating the essential motions in complex systems.[39,40] Figure 7A shows the relative contributions of different modes (eigenvectors) to the overall motion of the cyclin-CDK2 complexes. The figure clearly shows that the bulk of the protein dynamics can be


dx.doi.org/10.1021/bi5004052 | *Biochemistry* 2014, 53, 4612−4624

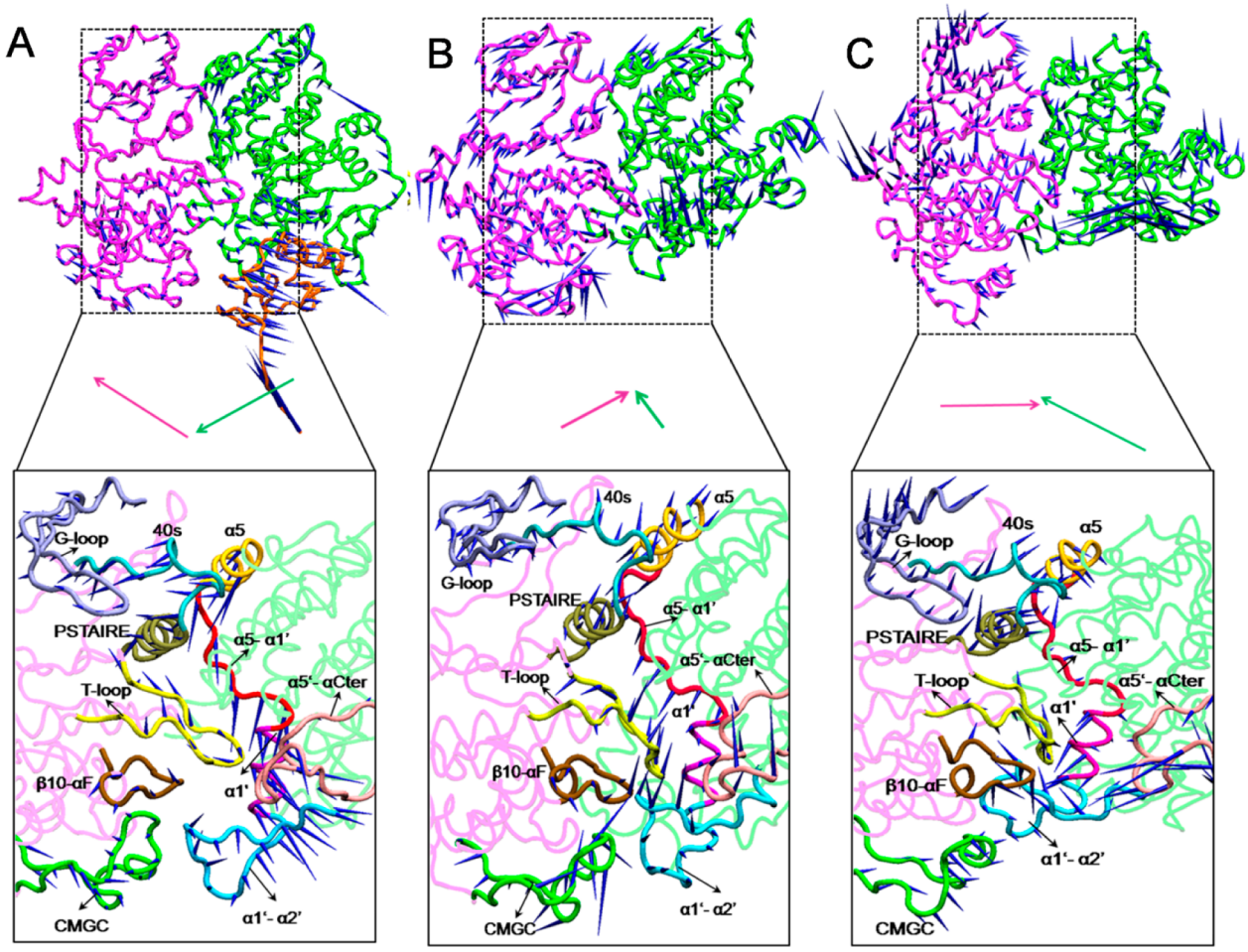



**Figure 8.** Porcupine plots representing the principal motions of cyclin E (green) and CDK2 (pink) along the direction of PC1. Results are shown for (A) full-length, (B) Trunk 1, and (C) Trunk 2 complexes. Directions and magnitudes of the net eigenvectors from cyclin E and CDK2 are shown in the inset. Secondary structural elements that exhibited most significant movements are enlarged and shown in different colors: cyclin E, helix $\alpha$5, orange; loop $\alpha$5-$\alpha$1′, red; helix $\alpha$1′, magenta; loop $\alpha$1′-$\alpha$2′, bright cyan; and loop $\alpha$5′-$\alpha$C-ter, pink. CDK2-T-loop, yellow; PSTAIRE helix, tan; loop $\beta$10-$\alpha$F, brown; CMGC domain, green; G-loop, ice blue; and 40s loop, cyan.

described by a small number of eigenvectors, which represent collective motions of the atoms. The first five eigenvectors can describe about 80% of the total mean square fluctuations of the full-length complex and about 70% of that of the Trunk 1 and Trunk 2 complexes. In order to compare regions sampled by the essential motions in different complexes, 2D projections of MD ensembles onto the plane defined by the first two eigenvectors were plotted (Figure 7B,C). Notable differences in fluctuation were observed among the three complexes. As these figures show, although a considerable overlap in the cluster of conformations was observed, the dynamics of truncated complexes is more restricted compared to that of the full-length complex. This is consistent with the *B*-factor data in Figure S6 (Supporting Information). As expected, CDK2 exhibits almost similar dynamics in all three complexes (Figure 7B).

The eigenvectors can be better interpreted by visualizing them through the porcupine plots.[39] To visualize an eigenvector, a cone is drawn for each residue starting from its C$\alpha$ and projecting in the direction of the component of the same eigenvector that corresponds to that residue. The length of the cone represents the magnitude of the motion. Figure 8 shows the principal motions of cyclin E and CDK2 residues in three complexes, along the direction of principal component 1 (PC1). The cyclin E residues that are known to make the primary contacts with CDK2, such as loops $\alpha$1′-$\alpha$2′, $\alpha$5- $\alpha$1′, and $\alpha$5′-$\alpha$Cter, and helices $\alpha$1′ and $\alpha$5, show the largest average velocity covariance vectors. Moreover, remarkable concerted motions are seen between these cyclin E elements and the interface elements of the CDK2:T-loop, PSTAIRE helix, 40s loop, etc. Even more importantly, PC1 showed a significant difference in the direction of displacements of these elements in the three complexes. While the displacements are approximately along the radial direction, with cyclin E and CDK2 vectors directing to each other in the truncated complexes; the movements and direction of the vectors are largely random in the full-length complex. The resultant CDK and cyclin eigenvectors from PC1, shown in the inset of Figure 8 also show a similar trend, where the directions of cyclin and CDK residue fluctuations in truncated complexes appear to complement each other, unlike in the full-length complex. These radial and complementary movements of the interface residues of cyclin E and CDK2 might imply better interactions and, therefore, are likely to facilitate complex formation. On the contrary, the random movements of both cyclin E and CDK2 residues may not allow efficient complex formation and could produce a less active complex. To recall, the pCDK2/cyclin E1 interface is the key to activating properties of the cyclins to the


dx.doi.org/10.1021/bi5004052 | *Biochemistry* 2014, 53, 4612−4624

CDKs. Interestingly, the substrate binding pocket and G-loop also show a larger concerted motion in the truncated complexes, where the velocity covariance vectors were longer and protruding outward, implying better substrate binding capacity of the truncated complexes.

## SUMMARY AND CONCLUSIONS

The crystal structure of the cyclin E1/CDK2 complex has been solved only very recently. The structure contains a truncated version of cyclin E that has both N- and C-terminal domains missing. *In vivo* and *in vitro* studies have shown that these truncated isoforms of cyclin E activate CDK2 more robustly than the full-length cyclin E. However, the molecular basis of higher activity of the truncated cyclin E/CDK2 complexes was unknown.

Our detailed bioinformatics analyses and molecular dynamics simulation results show that the missing regions in LMW isoforms of cyclin E contain NLS and PEST domains and are intrinsically disordered. These missing regions are rich in hydrophilic residues, have no or minimal secondary structures, and possess significantly higher flexibility than the crystallized part. However, while bound to the crystallized part in the full-length complex, these regions show an increased tendency to form secondary structures. They also follow the important prerequisite for the absence of regular structure: the combination of low mean hydrophobicity and relatively large net charge. These findings of disorderness of the terminal regions are consistent with the fact that (i) NLS and PEST domains in many proteins are reported to be intrinsically disordered and that (ii) only the central region of cyclin E, which is globular in structure, could be crystallized, while the terminal regions could not be captured in the electron density map.

These intrinsically disordered cyclin E1 regions weaken the cyclin E1/CDK2 interface, as evident from the relatively smaller contacts and larger solvent accessibility of the interacting residues in the full-length complex. The key interactions that are found in the crystal structure and simulated truncated complexes, e.g. van der Waals interactions between cyclin E:Val250$_{\alpha1'\text{-}\alpha2'}$, Leu251$_{\alpha1'\text{-}\alpha2'}$ and CDK2:Phe152$_{\text{T-loop}}$, Val156$_{\text{T-loop}}$ and also between cyclin E:Val237$_{\alpha1'}$ and CDK2:Val154$_{\text{T-loop}}$; hydrogen bonding interactions between main chains of cyclin E:Leu251$_{\alpha1'\text{-}\alpha2'}$ and CDK2:Val156$_{\text{T-loop}}$; stacking interactions between cyclin E:Pro253$_{\alpha1'\text{-}\alpha2'}$ and CDK2:Tyr179$_{\text{T-loop-}\alpha\text{F}}$; electrostatic interactions between cyclin E:Arg225$_{\alpha5\text{-}\alpha1'}$ and CDK2:Glu57$_{\text{PSTAIRE}}$ and also between cyclin E:Glu340$_{\alpha5'\text{-}\alpha\text{Cter}}$ and CDK2:Lys178$_{\text{T-loop-}\alpha\text{F}}$; and hydrogen bonding network involving cyclin E:Asp341$_{\alpha5'\text{-}\alpha\text{Cter}}$, Asn344$_{\alpha5'\text{-}\alpha\text{Cter}}$ and CDK2:Arg157$_{\text{T-loop}}$, were lost or became weak in the full-length complex.

The above findings were complemented by the principal motions of the two proteins in truncated complexes, where the interfacial residues of both the proteins exhibited large concerted motions along the radial direction to facilitate complex formation. On the contrary, the random movements of cyclin E residues in the full-length complex could not allow an efficient complex formation and produced a less active complex. To summarize, the cyclin E LMW isoforms assist the formation of a stronger protein−protein complex and, therefore, can induce G1-to-S phase progression in cancer cells, where they are overexpressed.

Finally, the results are validated by creating a model truncated complex of cyclin E (81−363)/CDK2 from the built structure of the full-length complex and simulating it for 50 ns. The final simulated structure was found to be very similar to the crystal structure of cyclin E (81−363)/CDK2, with correct reproduction of the interface residues and their contacts and interaction energies. Nevertheless, this study paves the way to conceive experiments for further validating our findings. For example, future work can attempt to confirm structural disorderness in missing N- and C-terminal domains by the far-UV CD spectroscopic technique[49,59] and Raman optical activity spectra.[60] More advanced techniques like laser temperature jump spectroscopy can be used to study the folding kinetics of these domains and their interactions with truncated CDK2/cyclin E complex.[61]

## ASSOCIATED CONTENT

### Supporting Information

Root mean squared displacements of full-length cyclin E/CDK2 complex and its further comparison with the truncated complexes. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION

### Corresponding Author

*Phone: +91-44-22574122. Fax: +91-44-22574102. E-mail: sanjibs@iitm.ac.in.

### Funding

We acknowledge Department of Biotechnology (DBT), Government of India for financial support.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Mr. Ajit Prabhakar for helping us with the initial model building of the full-length cyclin E. The computer resources of Computer Centre, IIT Madras are gratefully acknowledged.

## ABBREVIATIONS

CDK, cyclin dependent kinase; LMW, low molecular weight; MD, molecular dynamics; CAK, cyclin activating kinase; NLS, nuclear localization sequence; PEST, peptide sequence that is rich in proline (P), glutamic acid (E), serine (S), and threonine (T); IDP, intrinsically disordered protein; IDR, intrinsically disordered region; C−H, charge hydrophobicity; PCA, principal component analysis

## REFERENCES

(1) Morgan, D. O. (1997) Cyclin-dependent kinases: Engines, clocks, and microprocessors. *Annu. Rev. Cell Dev. Biol. 13*, 261−291.

(2) Kannan, N., and Neuwald, A. F. (2004) Evolutionary constraints associated with functional specificity of the CMGC protein kinases MAPK. *Protein Sci. 13*, 2059−2077.

(3) Malumbres, M. (2011) Physiological relevance of cell cycle kinases. *Physiol. Rev. 91*, 973−1007.

(4) Malumbres, M., and Barbacid, M. (2009) Cell cycle, CDKs and cancer: a changing paradigm. *Nat. Rev. Cancer 9*, 153−166.

(5) Pavletich, N. P. (1999) Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. *J. Mol. Biol. 287*, 821−828.

(6) Radzio-Andzelm, E., Lew, J., and Taylor, S. (1995) Bound to activate: conformational consequences of cyclin binding to CDK2. *Structure 3*, 1135−1141.
(7) Russo, A. A., Jeffery, P. D., and Pavletich, N. P. (1996) Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat. Struct. Biol. 3*, 696−700.
(8) De Bondt, H. L., Rosenblatt, J., Jancarik, J., Jones, H. D., Morgan, D. O., and Kim, S. H. (1993) Crystal structure of cyclin-dependent kinase 2. *Nature 363*, 595−602.
(9) Jeffery, P. D., Russo, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., and Pavletich, N. P. (1995) Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature 376*, 313−320.
(10) Möröy, T., and Geisen, C. (2004) Cyclin E. *Int. J. Biochem. Cell Biol. 36*, 1424−1439.
(11) Hwang, H. C., and Clurman, B. E. (2005) Cyclin E in normal and neoplastic cell cycles. *Oncogene 24*, 2776−2786.
(12) Honda, R., Lowe, E. D., Dubinina, E., Skamnaki, V., Cook, A., Brown, N. R., and Johnson, L. N. (2005) The structure of cyclin E1/CDK2: implications for CDK2 activation and CDK2-independent roles. *EMBO J. 24*, 452−463.
(13) Moore, J. D., Kornbluth, S., and Hunt, T. (2002) Identification of the nuclear localization signal in xenopus cyclin E and analysis of its role in replication and mitosis. *Mol. Biol. Cell 13*, 4388−4400.
(14) Jackman, M., Kubota, Y., Elzen, N. Den., and Hagting, A. (2002) Cyclin A- and cyclin E-Cdk complexes shuttle between the nucleus and the cytoplasm. *Mol. Biol. Cell 13*, 1030−1045.
(15) Rechsteiner, M., and Rogers, S. W. (1996) PEST sequences and regulation by proteolysis. *Trends Biochem. Sci. 21*, 267−271.
(16) Gudas, J. M., Payton, M., Thukral, S., Chen, E., Bass, M., Robinson, M. O., and Coats, S. (1999) Cyclin E2, a novel G1 cyclin that binds Cdk2 and is aberrantly expressed in human cancers. *Mol. Cell. Biol. 19*, 612−622.
(17) Harwell, R. M., Mull, B. B., Porter, D. C., and Keyomarsi, K. (2004) Activation of cyclin-dependent kinase 2 by full length and low molecular weight forms of cyclin E in breast cancer cells. *J. Biol. Chem. 279*, 12695−12705.
(18) Harwell, R. M., Porter, D. C., Danes, C., and Keyomarsi, K. (2000) Processing of cyclin E differs between normal and tumor breast cells. *Cancer Res. 60*, 481−489.
(19) Wingate, H., Bedrosian, I., Akli, S., and Keyomarsi, K. (2003) The low molecular weight (LMW) isoforms of cyclin E deregulate the cell cycle of mammary epithelial cells. *Cell Cycle 2*, 461−466.
(20) Spruck, C., Sun, D., Fiegl, H., Marth, C., Mueller-Holzner, E., Goebel, G., Widschwendter, M., and Reed, S. I. (2006) Detection of low molecular weight derivatives of cyclin E1 is a function of cyclin E1 protein levels in breast cancer. *Cancer Res. 66*, 7355−7360.
(21) Geisen, C., and Moroy, T. (2002) The oncogenic activity of cyclin E is not confined to Cdk2 activation alone but relies on several other, distinct functions of the protein. *J. Biol. Chem. 277*, 39909−39918.
(22) Porter, D. C., and Keyomarsi, K. (2000) Novel splice variants of cyclin E with altered substrate specificity. *Nucleic Acids Res. 28*, 1−8.
(23) Porter, D. C., Zhang, N., Danes, C., Gahren, M. J. M. C., Harwell, R. M., Faruki, S., and Keyomarsi, K. (2001) Tumor-specific proteolytic processing of cyclin E generates hyperactive lower-molecular-weight forms. *Mol. Cell. Biol. 21*, 6254−6269.
(24) UniProt Consortium. (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res. 38*, D142−148.
(25) Xu, D., and Zhang, Y. (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins 80*, 1715−1735.
(26) Xu, D., and Zhang, Y. (2013) Towards optimal fragment generations for ab initio protein structure assembly. *Proteins 81*, 229−239.
(27) Eswar, N., Marti-Renom, M. A., Webb, B., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U., and Sali, A. (2006) Comparative protein structure modeling with MODELLER. *Curr. Protoc. Bioinf. 15*, 5.6.1−5.6.30.
(28) Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283−291.
(29) Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta. Crystallogr., Sect. D 66*, 12−21.
(30) Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., et al. (2012) *AMBER 12*, University of California, San Francisco, CA.
(31) Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graphics 8*, 52−56.
(32) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926−935.
(33) Meagher, K. L., Redman, L. T., and Carlson, H. A. (2003) Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem. 24*, 1016−1025.
(34) Bradbrook, G. M., Gleichmann, T., Harrop, S. J., Habash, J., Raftery, J., Kalb, J., Yariv, J., Hillier, I. H., and Helliwell, J. R. (1998) X-ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes - Relating structure to thermodynamics of binding. *J. Chem. Soc. Faraday. Trans. 94*, 1603−1611.
(35) Craft, J. W., and Legge, G. B. (2005) An AMBER/DYANA/MOLMOL phosphorylated amino acid library set and incorporation into NMR structure calculations. *J. Biomol. NMR. 33*, 15−24.
(36) Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Lee, G. P. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys. 103*, 8577−8593.
(37) Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys. 23*, 327−341.
(38) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins 65*, 712−725.
(39) Natarajan, K., and Senapati, S. (2013) Probing the conformational flexibility of monomeric FtsZ in GTP-bound, GDP-bound, and nucleotide-free states. *Biochemistry 52*, 3543−3551.
(40) Berendsen, H. J. C., and Hayward, S. (2000) Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol. 10*, 165−169.
(41) Meyer, T., Ferrer-Costa, C., Perez, A., Rueda, M., Bidon-Chanal, A., Luque, F. J., Laughton, C. A., and Orozco, M. (2006) Essential dynamics: A tool for efficient trajectory compression and management. *J. Chem. Theory Comput. 2*, 251−258.
(42) Hiltpold, A., Ferrara, P., Gsponer, J., and Caflisch, A. (2000) Free energy surface of the helical peptide Y$(MEARA)_6$. *J. Phys.Chem. B 104*, 10080−10086.
(43) Freddolino, P. L., and Schulten, K. (2009) Common structural transitions in explicit-solvent simulations of villin headpiece folding. *Biophys. J. 97*, 2338−2347.
(44) Senapati, S., Bui, J. M., and McCammon, J. A. (2005) Induced fit in mouse acetylcholinesterase upon binding a femtomolar inhibitor: a molecular dynamics study. *J. Med. Chem. 48*, 8155−8162.
(45) Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins 41*, 415−427.
(46) Uversky, V. N. (2002) What does it mean to be natively unfolded? *Eur. J. Biochem. 269*, 2−12.
(47) Dunker, K. A., Silman, I., Uversky, V. N., and Sussman, J. L. (2008) Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol. 18*, 756−764.
(48) Baker, E. S., Luckner, S. R., Krause, K. L., Lambden, P. R., Clarke, I. N., and Ward, V. K. (2012) Inherent structural disorder and dimerisation of murine norovirus NS1−2 protein. *PLoS One 7*, 1−13.
(49) Maestro, B., Galán, B., Alfonso, C., Rivas, G., Prieto, M. A., and Sanz, J. M. (2013) A new family of intrinsically disordered proteins:

structural characterization of the major phasin PhaF from *Pseudomonas putida* KT2440. *PLoS One 8*, 1−15.
(50) Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol. 157*, 105−132.
(51) Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., and Uversky, V. N. (2010) PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta 1804*, 996−1010.
(52) McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics 16*, 404−405.
(53) Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci. 11*, 739−756.
(54) Demarest, S. J., Zhou, S. Q., Robblee, J., Fairman, R., Chu, B., and Raleigh, D. P. (2001) A comparative study of peptide models of the alpha-domain of alpha-lactalbumin, lysozyme, and alpha-lactalbumin/lysozyme chimeras allows the elucidation of critical factors that contribute to the ability to form stable partially folded states. *Biochemistry 40*, 2138−2147.
(55) Espinoza-Fonseca, L. M. (2009) Reconciling Binding mechanisms of intrinsically disordered proteins. *Biochem. Biophys. Res. Commun. 382*, 479−482.
(56) 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.
(57) Wright, P. E., and Dyson, H. J. (2009) Linking folding and binding. *Curr. Opin. Struct. Biol. 19*, 31−38.
(58) Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E., and Edelman, M. (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics 15*, 327−332.
(59) Jobby, M. K., and Sharma, Y. (2007) Caulollins from Caulobacter crescentus, a pair of partially unstructured proteins of beta gamma-Crystallin superfamily, gain structure upon binding calcium. *Biochemistry 46*, 12298−12307.
(60) Smyth, E., Syme, C. D., Blanch, E. W., Hecht, L., Vasák, M., and Barron, L. D. (2001) Solution structure of native proteins with irregular folds from Raman optical activity. *Biopolymers 58*, 138−51.
(61) Narayanan, R., Ganesh, O. K., Edison, A. S., and Hagen, S. J. (2008) Kinetics of folding and binding of an intrinsically disordered protein: the inhibitor of yeast aspartic proteinase YPrA. *J. Am. Chem. Soc. 130*, 11477−11485.