PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

The flexible C-terminal arm of the Lassa arenavirus Z-protein mediates interactions with multiple binding partners

Eric R. May, Roger S. Armen, Aristotle M. Mannan, and Charles L. Brooks III*

Department of Chemistry and Program in Biophysics, University of Michigan, Ann Arbor, Michigan 48109

ABSTRACT

The arenavirus genome encodes for a Z-protein, which contains a RING domain that coordinates two zinc ions, and has been identified as having several functional roles at various stages of the virus life cycle. Z-protein binds to multiple host proteins and has been directly implicated in the promotion of viral budding, repression of mRNA translation, and apoptosis of infected cells. Using homology models of the Z-protein from Lassa strain arenavirus, replica exchange molecular dynamics (MD) was used to refine the structures, which were then subsequently clustered. Population-weighted ensembles of low-energy cluster representatives were predicted based upon optimal agreement of the chemical shifts computed with the SPARTA program with the experimental NMR chemical shifts. A member of the refined ensemble was indentified to be a potential binder of budding factor Tsg101 based on its correspondence to the structure of the HIV-1 Gag late domain when bound to Tsg101. Members of these ensembles were docked against the crystal structure of human eIF4E translation initiation factor. Two plausible binding modes emerged based upon their agreement with experimental observation, favorable interaction energies and stability during MD trajectories. Mutations to Z are proposed that would either inhibit both binding mechanisms or selectively inhibit only one mode. The C-terminal domain conformation of the most populated member of the representative ensemble shielded protein-binding recognition motifs for Tsg101 and eIF4E and represents the most populated state free in solution. We propose that C-terminal flexibility is key for mediating the different functional states of the Z-protein.

Proteins 2010; 78:2251–2264.

Key words: chemical shifts; replica exchange molecular dynamics; SPARTA; RING domain; protein–protein docking.

INTRODUCTION

Lassa virus is a member of the arenavirus family and is distributed in Western Africa. Humans infected by some arenavirus strains, including Lassa, can develop a fatal case of hemorrhagic fever.[1,2] Arenavirus infection has recently been identified as an emerging health treat in the United States,[3–5] and there are currently no FDA approved vaccines or therapeutic agents.[6,7] Arenaviruses are segmented negative-strand RNA viruses and are enveloped by a lipid membrane. The genome codes for a glycoprotein that is posttranslationally cleaved into GP1 and GP2, a nucleoprotein (NP), a viral polymerase (L), and a zinc-binding protein (Z). No structure of any Lassa protein has been determined; no structures of any other arenavirus protein have been determined except for GP1 of the Machupo virus.[8] The Z-protein of Lassa, and other arenavirus strains, contains a RING domain that coordinates two zinc ions. The role of Z in the life cycle of arenaviruses is not well understood, but Z is believed to have a structural role because of its known interactions with both the NP[9] and the envelope membrane through N-terminal myristoylation.[10] However, Z also interacts with several host proteins, indicating it has several functional roles as well. Z interacts with protein Tsg101 to promote virus budding from cells,[11] analogous to HIV-1 Gag[12] and Ebola matrix protein VP40.[13] Colocalization of Z and the human oncoprotein promyelocytic leukemia protein (PML), which also contains a RING domain, has been observed in cells infected by the arenavirus lymphocytic choriomeningitis virus.[14] The presence of Z acts to redistribute PML nuclear bodies to the cytoplasm, and this interaction is suggested to have implications for arenavirus virulence and in acute promyelocytic leukemia, which reduces cell's ability to perform apo-

Abbreviations: CS, chemical shifts; CS-rmsd, chemical shift root mean squared deviation; fs, femtosecond; ns, nanosecond; REX, replica exchange molecular dynamics; rmsd, root mean squared deviation.

The PBD structures of R^{Full} and the best static and best dynamic complexes of eIF4E:Z are available from the corresponding author upon request.

Grant sponsor: National Institutes of Health; Grant numbers: RR012255; Grant number: RR01081; Grant sponsor: National Science Foundation (Postdoctoral Fellowship); Grant number: DBI-0905773.

Eric R. May and Roger S. Armen contributed equally to this work.

*Correspondence to: Charles L. Brooks III, Department of Chemistry and Program in Biophysics, 930 N. University Ave, University of Michigan, Ann Arbor, MI 48109. E-mail: brookscl@umich.edu.

Received 21 January 2010; Revised 7 March 2010; Accepted 20 March 2010
Published online 5 April 2010 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22738

ptosis. Other interaction partners of Z include proline-rich homeodomain protein,[15] ribosomal protein P0,[16] and eukaryotic initiation factor eIF4E.[17] eIF4E plays an important role in the regulation of translation initiation, and when Z binds to eIF4E it represses translation, which contributes to the characteristic chronic infection of arenavirus. The ability of Z to interact with several proteins, and the parallel to other dangerous viruses, makes structural characterization of Z an important step forward in understanding arenavirus infection.

Recently, the NMR assignment of Lassa Z-protein was published,[18] and the chemical shifts were deposited in the BMRB database (http://www.bmrb.wisc.edu/). The ^{15}N HSQC spectrum was assigned for all but 11 residues, indicating the structure is partially well defined in solution.[18] From the ^{1}H-^{15}N NOEs, it is clear that the most defined part of the structure is the RING domain (res: 26–70), and both termini are expected to be much more dynamic.[18] From the C^{α} chemical shifts an α-helix was clearly present between residues Leu51 and Leu57. In this work, we are interested in elucidating an ensemble of conformations of the Lassa Z-protein consistent with the limited structural information available and examining the structural consequences of its members in light of some of the functional roles of the Z-protein.

The notion of predicting a protein structure based upon chemical shift data is relatively novel. The recently introduced CHESHIRE method uses a fragment-based approach, which incorporates chemical shift data into the building, refinement, and scoring stages of structure prediction.[19] The ROSETTA program has been successful in the *de novo* structure prediction of small proteins from only amino acid sequence.[20] A new extension, CS-ROSETTA, generates fragments based on the CS, which are then used to build the full-length models with ROSETTA, and then scores the models accounting for their CS deviation from experimental values.[21,22] Both CHESHIRE and CS-ROSSETTA were validated in independent test sets and seemed to yield accurate predictions; both programs predicted all structures within their test sets to be within 2.0 Å backbone root mean squared deviation (rmsd) of the known structures. It should be noted that both test sets consisted of well-structured proteins with a high percentage of α-helix and/or β-sheet secondary structural elements. The CHESHIRE test set contained proteins with at least 51% secondary structure (α-helix and/or β-sheet), and all proteins in the CS-ROSETTA test set contained at least 47% secondary structure.

To predict the structure of the Lassa Z protein, we took an approach that combined elements of homology modeling, molecular dynamics (MD), and the available experimental data from NMR. We generated initial homology model structures using a multiple template method (I-TASSER)[23,24] and from single templates (MODELLER).[25] The initial models underwent refinement with the replica exchange MD protocol (REX),[26] which has been successful at refining structures with the incorporation of NMR data.[27,28] Several chemical shift prediction servers exist, including SHIFTS,[29,30] SHIFTX,[31] PROSHIFT,[32] CheShift,[33] and SPARTA.[34] In this work, the chemical shifts of the refined structures were predicted with the SPARTA program, which is also used by CS-ROSETTA, and were shown to outperform other CS calculation programs on a test set of 200 proteins.[34] The dynamic regions of Z and the lack of identified secondary structural elements (7% structured) may lead to a diverse conformational ensemble and indicate that a single structural model is less likely to agree with the experimental data. By analogy to NMR ensembles, we constructed weighted ensembles of energetically favorable structures to best match the experimental chemical shifts. The construction of weighted ensembles, optimized against NMR data, for unfolded protein states has been explored previously with the programs ENSEMBLE[35] and MINOES,[36] and our method follows in a similar vein. We then attempted to both validate and assign function to the ensemble members through a protein–protein docking study with the host protein eIF4E. The docking results were not only consistent with the experimental observations but also allowed us to glean biological insight and make connections between protein dynamics and the various functional roles of the Z protein.

RESULTS AND DISCUSSION

The only available experimentally determined structural information about the Lassa Z-protein is the CS, determined by NMR and deposited in the BMRB database (accession number 15660).[18] The only secondary structural feature determined, based upon the C^{α} shifts, was an α-helix present between residues Leu51 and Leu57.[18] This information was incorporated into our methodology for predicting the structure of the Z-protein.

In the subsections below, we begin by examining the unrefined homology models. We then examine the REX refined structures and detail the methodology we use to go from the initial homology models to our predicted ensembles. Several ensembles are constructed, beginning with the region that spans the two zinc centers (res: 31–67), which is expected to constrain the topology of the protein and be the most rigid segment; we will refer to this segment as the core. We extend our analysis to include the C-terminal segment and the core (res: 31–99) and then finally the full-length protein (res: 1–99). From the ensemble constructed for the full protein, we conduct further studies to explore the plausibility of ensemble members binding Z's known interaction partners Tsg101 and eIF4E.

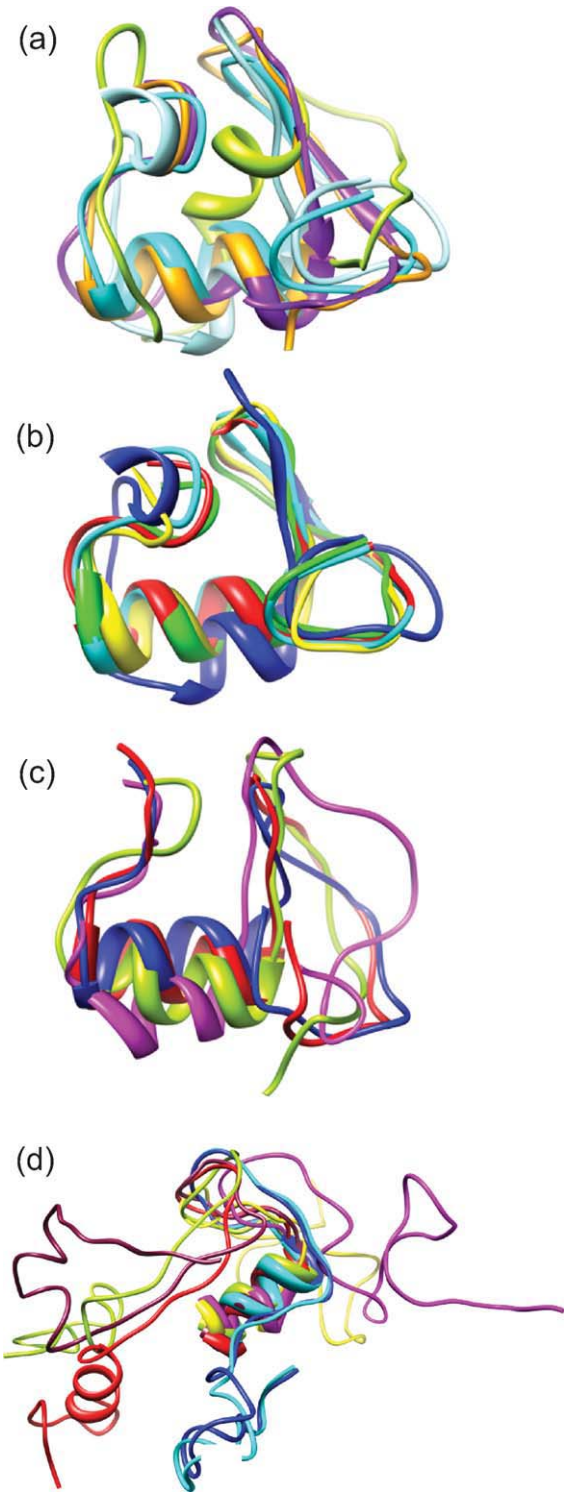

Figure 1

Comparison of models of the core and C-terminus. (**a**) Core configurations of model 1 for each of the initial homology models. (**b**) Core configurations of the five most populated members of $\boldsymbol{H}^{\text{Full}}$. (**c**) Core configuration of the four most populated members of the $\boldsymbol{R}^{\text{Core}}$. (**d**) C-terminus orientations of the seven most populated conformations of $\boldsymbol{R}^{31-99}$ (res: 51–99 are shown).

Homology models

Homology modeling yielded 25 acceptable structural models (20 from MODELLER and five from I-TASSER, as described in the Methods section), 17 of the models had helix in the predicted region. The core of the best model from I-TASSER and the best model from each of the four acceptable templates used by MODELLER are shown in Figure 1(a). For the five models shown in Figure 1(a), the average C^{α} rmsd over the core is 4.7 Å. As the experimental chemical shifts are determined from an ensemble of configurations, we were interested in constructing an ensemble that would optimally agree with the experimental data. The quantity we chose to minimize, in predicting an ensemble that best matched the experimental CS, was the chemical shift root mean squared deviation (CS-rmsd) calculated over all matching atoms between the experimental and predicted CS data sets. We first filtered out structures without helix in the predicted region, and with the remaining 17 structures a minimal CS-rmsd ensemble was constructed. The ensemble consisted of 10 structures (those with $w_j \geq 1.0\%$), and the CS-rmsd of the ensemble was 1.94 ppm; this will be referred to as the homology ensemble ($\boldsymbol{H}^{\text{Full}}$). The members of $\boldsymbol{H}^{\text{Full}}$ with the five largest weights contribute 80% of the population and are shown in Figure 1(b). The members of the $\boldsymbol{H}^{\text{Full}}$ display a consistent topology and similar tertiary structure within the core; however, both the C-terminal and N-terminal segments display large variations (not shown). The structural C^{α} rmsd was calculated between all 10 members of the $\boldsymbol{H}^{\text{Full}}$ and ranged from 0.3 to 3.6 Å over the core and 7.4 to 18.4 Å over the entire protein.

Replica exchange refined structures

By considering the final 101 structures from REX (at the lowest temperature) for all 25 homology models, 2525 structures were available for analysis. The clustering and subsequent energy evaluation can be performed on the entire protein or only on a segment of the protein. In our analysis of the REX structures, we took a hierarchical approach by first evaluating only the core, followed by the core and the C-terminal segment, which contains the late domains, and finally the entire protein including the N-terminal segment. When analyzing (energy evaluation, CS prediction, clustering) only a segment of the protein, the calculations were preformed by ignoring the rest of the protein. The terms representative structure and lowest energy structure will be used synonymously in referring to the structure with the lowest computed energy (from CHARMM) in each cluster.

Core segment: residues 31–67

The clustering tree of the core shows that there are two main clusters (see Fig. 2). The larger cluster consisted of 2020 members (Cluster 1) and the single lowest

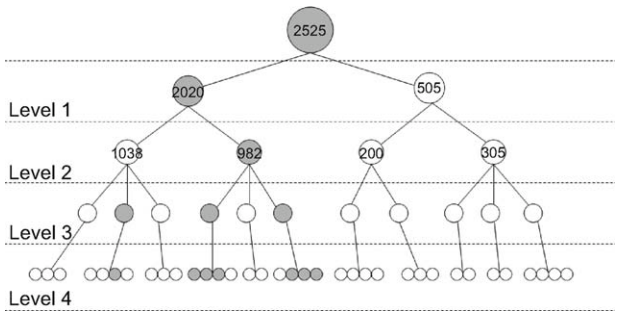

Figure 2

Hierarchal clustering of the core (res: 31–67). Only the first four (of eight) levels of clustering are shown. The numbers indicate the population size of the clusters. Clusters whose lowest energy structure had a helix from Leu51 to Leu57 are colored gray.

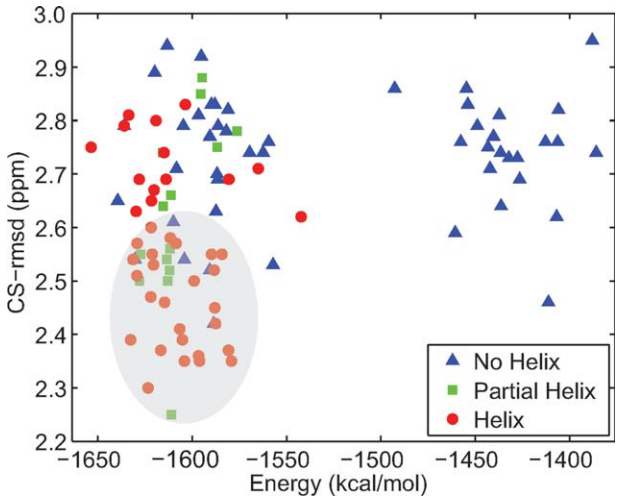

Figure 3

CS-rmsd versus energy for $\boldsymbol{R}^{\mathrm{Core}}$; data points are colored on the basis of helical content over residues Leu51–Leu57.

energy structure of the cluster has a helix spanning residues Leu51–Leu57. In Figure 2 the clusters are colored gray if the representative structure has helix spanning Leu51–Leu57. It can be observed that the smaller cluster at Level 1 (Cluster 2) does not contain any representative structures with helix spanning Leu51–Leu57, whereas Cluster 1 has representative structures with helix and others without helix. Cluster 2 consists mainly of structures that are mostly unfolded in the core; the representative structure of Cluster 2 has much higher energy (−1492.7 kcal/mol) than the representative structure of Cluster 1 (−1653.6 kcal/mol). When examining all the representative structures, shown in Figure 3, the low-energy structures display a wide range of CS-rmsd values, whereas the high-energy structures correspond to high CS-rmsd. Among those structures with both low energy and low CS-rmsd (gray region in Fig. 3), there is a prevalence of structures containing helix in the predicted region. To construct an ensemble of representative structures, we considered only structures with a helix spanning the predicted helical region (res: 51–57). We began by performing averaging starting with the three representative structures with helix at the third clustering level (those shown in gray at Level 3 of Fig. 2). The same procedure was preformed at all subsequent clustering levels, up to the terminal eighth level. It should be noted that there is redundancy of representative structures between cluster levels; this is due to the fact that a representative structure at a given level will always be present at subsequent levels, and branches that terminate before the terminal level are considered present at each level beyond their termination. The results of this configurational averaging are presented in Table I, and it can be seen that the averaged CS-rmsd decreases with increasing clustering level. From a pool of 40 structures at the terminal clustering level, whose individual CS-rmsds ranged from 2.30 to 2.83 ppm, the minimization algorithm constructs a seven-member ensemble whose weighted average CS-rmsd is 1.99 ppm; this will be termed the representative ensemble of the core ($\boldsymbol{R}^{\mathrm{Core}}$). The ensemble averaged CS-rmsd is considerably lower than the best individual CS-rmsd of the entire pool of 2525 structures (2.11 ppm). From $\boldsymbol{R}^{\mathrm{Core}}$, the four most dominant members (which contribute 85% based upon the w_js) are shown in Figure 1(c). There is a consistent topology among all members of the $\boldsymbol{R}^{\mathrm{Core}}$. The structural C^{α} rmsd between the members of $\boldsymbol{R}^{\mathrm{Core}}$ ranged from 2.4 to 5.7 Å with a population-weighted average of 3.7 Å. The structural rmsd information is presented in Table II for $\boldsymbol{R}^{\mathrm{Core}}$, $\boldsymbol{H}^{\mathrm{Full}}$, and the other ensembles defined later in this work. The methodology of constructing the representative ensembles is described in the workflow diagram in Figure 4.

Core segment and C-terminus: residues 31–99

In trying to build toward a complete model of the Z-protein, we next considered the segment spanning the zinc centers to the C-terminus (res: 31–99). The C-terminal region of the protein is of particular interest because of the two proline-rich "late domains" it possesses

Table I

Calculated CS-rmsd Results Based upon Configurational Averaging

Clustering level	Initial # of structures	CS-rmsd range (ppm)	# $w_j > 0$	Averaged CS-rmsd (ppm)
3	3	2.39–2.76	3	2.29
4	7	2.30–2.76	3	2.10
5	17	2.30–2.67	4	2.02
6	30	2.30–2.81	5	2.00
7	37	2.30–2.81	7	1.99
8	40	2.30–2.83	7	1.99

Table II
Structural C^{α} rmsd Calculations for the Various Ensembles, Reported in Å

		rmsd (res: 31–67)		rmsd (res: 31–99)		rmsd (res: 1–99)	
Ensemble	Members	Range	Average	Range	Average	Range	Average
R^{Core}	7	2.4–5.7	3.7	n/a	n/a	n/a	n/a
$R^{31–99}$	10	0.9–5.6	3.9	1.2–15.1	10.7	n/a	n/a
$R^{Ful/}$	13	0.5–6.0	3.8	0.7–15.6	10.2	2.4–20.1	14.1
$H^{Ful/}$	10	0.3–3.6	2.3	2.0–15.5	10.5	7.4–18.4	13.5

The averages are population-weighted averages.

(PTAP, PPPY), which are implicated in viral envelope fission from the cellular membrane.[11,37] The PTAP motif is known to interact with Tsg101,[11] and therefore it is possible that the C-terminus is more structured than the N-terminus (or may undergo a structured to unstructured transition). We followed the same workflow (see Fig. 4) to construct the representative ensemble for this segment ($\boldsymbol{R}^{31–99}$) as was done the core.

When clustering is based upon residues 31–99, three clusters are formed at Level 1. At the terminal clustering level (Level 7), there are 35 representative structures that have the helix; the CS-rmsd values of these 35 structures range from 2.33 to 2.67 ppm. The minimization procedure results in the 10-member ensemble, $\boldsymbol{R}^{31–99}$, with an optimized CS-rmsd of 1.98 ppm. $\boldsymbol{R}^{31–99}$ displays the same core topology and basic tertiary structure as $\boldsymbol{R}^{Core}$. However, the C-terminus does not display a distinct conformation; multiple conformations are observed within $\boldsymbol{R}^{31–99}$. The seven most populated structures of $\boldsymbol{R}^{31–99}$ account for 93% of the population and are shown in Figure 1(d); three differing orientations of the C-terminus can be observed.

Full protein: residues 1–99

When clustering is based upon the entire protein sequence, a much broader distribution is observed compared with the previous two clustering schemes. At the first level, there are two clusters but at the second level there are 124 clusters; clustering from residues 31–67 and 31–99 returned four and seven clusters at Level 2, respectively. Level 4 was the terminal clustering level for the full protein, compared with terminal levels of 8 and 7 for the previous schemes. The interpretation of the broadness and shallowness of the clustering tree is that there is a greater diversity of configurations, which implies that the N-terminus is more flexible than the core or the C-terminus. Following the same methodology (see Fig. 4), the representative ensemble of the full protein ($\boldsymbol{R}^{Full}$) consisted of 13 structures producing a CS-rmsd of 1.91 ppm, which is lower than the CS-rmsd of $\boldsymbol{H}^{Full}$ (1.94 ppm). The core segment of $\boldsymbol{R}^{Full}$ displays the same topology and tertiary structure that was observed in $\boldsymbol{R}^{Core}$ and $\boldsymbol{R}^{31–99}$. The C-terminus orientations are similar to those of $\boldsymbol{R}^{31–99}$ (see Fig. 4). The N-terminus orientations within $\boldsymbol{R}^{Full}$ are widely varying, which is indicative of the flexibility of the region. We also observe fewer contacts between the N-terminus and the core when compared with the C-terminus with the core. Having the N-terminus free and flexible is likely related to the fact that in the mature viral particle the N-terminus is myristoylated[10] and inserted into the envelope membrane. It presumably serves as a flexible linker between the membrane and the core of Z, which may be bound to the major capsid protein and/or the GP spike.

We were interested to understand the relationship between $\boldsymbol{R}^{Full}$ and the original homology models (parent structure). Five members of $\boldsymbol{R}^{Full}$ originated from an I-TASSER model and these account for 56% of the population weight. We observe many instances when a structure has moved away from its parent structure and closer to a different homology model. The general observation is that REX shifted the structures significantly from their parent structure. Even over the most structured core region, the ensemble members moved up to 4.6 Å (C^{α} rmsd) away from their parent structure during REX refinement. Ultimately, this diversification of configurations allowed for better agreement with the experimental CS.

Effect of zinc ions on chemical shift predictions and ensemble construction

We note two potential sources of error in the CS calculation and resultant CS-rmsd minimized ensembles. The first is the uncertainty in the SPARTA calculations themselves, the standard deviations of the predicted

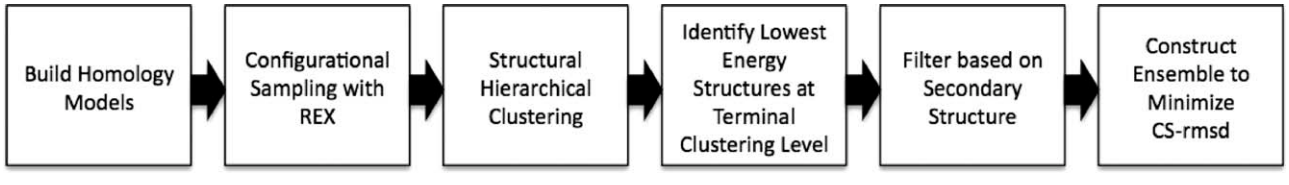

Figure 4
Workflow diagram describing the methodology of weighted ensemble construction.

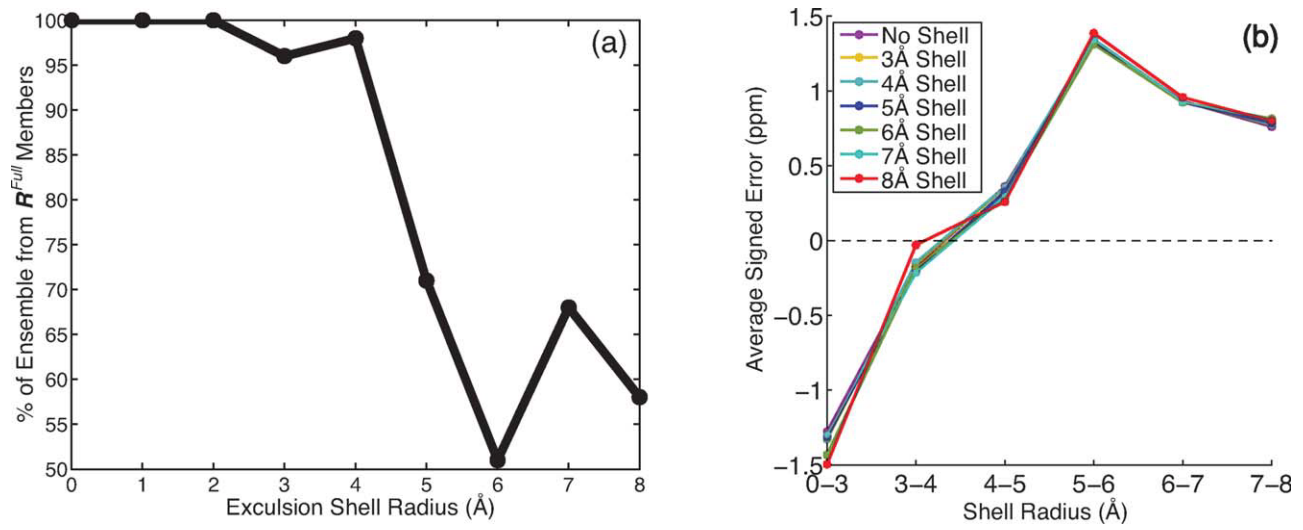

Figure 5

Evaluating the effect of zinc ions on ensemble construction by excluding atoms around the zincs from the CS calculations. (**a**) The percentage of the ensemble made from members of $\boldsymbol{R}^{\mathrm{Full}}$ versus the size of the exclusion shell. (**b**) The averaged signed CS error of atoms within a specified distance from the zincs for $\boldsymbol{R}^{\mathrm{Full}}$ (no shell) and the ensembles with varying exclusion shells.

chemical shifts range from 2.52 ppm for ^{15}N down to 0.27 ppm for $^{1}H^{\alpha}$, based upon their validation study.[34] The second, and potentially more significant, source of error arises from the neglect of the zinc ions in the CS predictions. The effect of metal ions on chemical shifts is not well understood, and there is no CS prediction program capable of accounting for these metal ions.

In an attempt to remove the influence of zinc from our ensemble construction, we removed those atoms closest to the zinc ions from the CS-rmsd optimization procedure. There are 467 atoms common to both the experimentally and computationally determined CS sets. From the pool of 61 structures used to build $\boldsymbol{R}^{\mathrm{Full}}$, we identified any atom, common to both CS sets, that fell within 3 Å of either zinc ion (the exclusion shell), removed them from the CS-rmsd optimization and constructed a new ensemble. We preformed this same analysis for increasing radius of the exclusion shell up to 8 Å. Figure 5(a) shows how much of the population of the ensembles, constructed with varying size exclusion shells, are from structures also present in $\boldsymbol{R}^{\mathrm{Full}}$. It can be seen that for small exclusion shells (up to 4 Å), the ensembles are nearly identical to $\boldsymbol{R}^{\mathrm{Full}}$, and even at an 8 Å exclusion shell, the majority of the ensemble derives from $\boldsymbol{R}^{\mathrm{Full}}$. It is known that the zinc ions should have a deshielding (downfield) effect,[38] leading to increased CS values in their vicinity. As this effect is not captured by the SPARTA CS predictions, one would expect the CS predictions on the atoms around zinc to be lower than the experimental values. For $\boldsymbol{R}^{\mathrm{Full}}$ and the ensembles constructed with exclusion shells, the averaged signed error ($CS^{Comp} - CS^{Exp}$) in concentric shells around the zinc ions was computed. This data are shown in Figure 5(b), and it can be seen that the error in a shell from 0 to 3 Å and 3 to 4 Å is negative, whereas all other shells show a positive error. Therefore, on an average, the predicted CS within 4 Å of a zinc ion is underestimated (upfield shifted), presumably because of the absence of zinc in the calculation. Beyond 4 Å from the zinc ions, on an average, the predicted CS is overestimated, and therefore this error is not dominated by the absence of zinc. On the basis of this evidence, we believe the zincs ions are causing a local (~4 Å) perturbation to the CS and are not a global influence. This seems reasonable as the majority of the zinc charge is being felt by the ligating side chains, which have lengths on the order 5 Å. Furthermore, when the atoms within 4 Å of the zinc ions are excluded from the CS prediction, an ensemble is constructed with 98% of its population common to structures in $\boldsymbol{R}^{\mathrm{Full}}$.

Conformation of the PTAP late domain on the C-terminal arm

The Z-protein displays a PTAP motif (also known as the "late domain") on the C-terminal region (res: 81–84), which is the recognition motif for binding to the human protein Tsg101 that recruits several other cellular factors that promote viral budding. The major capsid proteins of HIV-1 (Gag) and Ebola virus (matrix protein VP40) both display a PTAP late domain that binds to Tsg101 to facilitate viral budding. The crystal structure of a minimal nine-residue peptide of the HIV-1 Gag late domain bound to Tsg101 shows that the PTAP motif

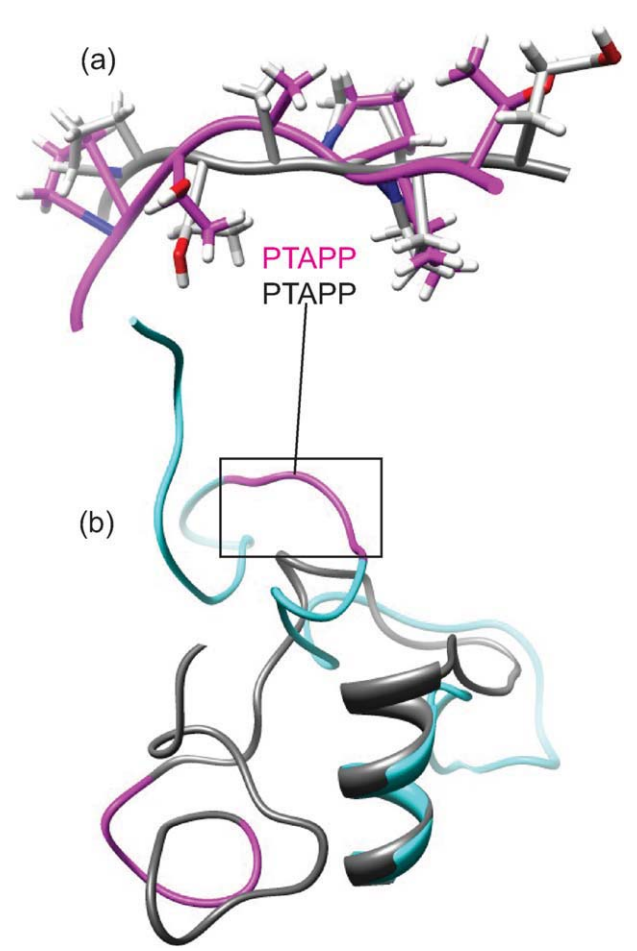

Figure 6

Conformation of the late domain (PTAPP) that is favorable for binding to Tsg101. (**a**) Superposition of PTAPPs from Z (magenta) and HIV-1 Gag (gray) bound to Tsg101 (1M4P). (b) The second most populated member of $\boldsymbol{R}^{\text{Full}}$ (cyan) presents a conformation of the late domain (magenta) that is favorable for binding Tsg101. The most populated member of $\boldsymbol{R}^{\text{Full}}$ is shown in gray. When these two conformations are superimposed over the core, a distance of 25 Å separates the PTAPP motifs. The gray conformation has extensive contacts between the C-terminus and the core, whereas the cyan conformation has very few.

binds in a partial left-handed type II polyproline helical conformation. It is plausible to expect the late domain of Z binds with a similar binding mode as HIV-1 Gag. We examined the PTAP late domain of $\boldsymbol{R}^{\text{Full}}$ to see if a similar conformation to the bound HIV-1 Gag existed. We found that the late domain adopted differing secondary structures including extended and partially helical conformations in our ensemble. The tertiary structure of some configurations had direct interactions between PTAP and the core of Z, whereas in other structures PTAP was free of any contacts with the core, allowing the C-terminus to sample a variety of conformational states. As the sequence of Z and the bound HIV-1 peptide share the five residues sequence PTAPP, we calculated the C^{α} rmsd over PTAPP of Z compared with the bound conformation of the HIV-1 Gag peptide.[39] The closest PTAPP conformation to HIV-1 Gag peptide was 0.9 Å C^{α} rmsd [Fig. 6(a)]. This conformation of Z had a C-terminal conformation with few contacts to the core, therefore presenting a tertiary structure and extended secondary structure capable of binding to Tsg101.

The structure we propose to be favorable for binding Tsg101 has the second largest population weight in $\boldsymbol{R}^{\text{Full}}$. When compared with the most populated configuration in $\boldsymbol{R}^{\text{Full}}$, many differences are observed in the C-terminal configurations [Fig. 6(b)]. In the most populated conformation, Pro85 has a favorable hydrophobic contact with Leu51 on the α-helix, which should partially order the C-terminus. This C-terminal conformation is additionally stabilized by favorable hydrophobic contacts between Ile92 and Leu56 on the α-helix as well as side-chain hydrogen bonds between Ser78-Thr55 and Ser91-Asn52. Because of these extensive contacts between the C-terminus and the core, we propose that the C-terminus and the PTAP late domain will undergo a structural transition from this partially ordered state to a more disordered state that would allow binding of Tsg101.

Predicting the structure of the eIF4E:Z complex with protein–protein docking

Translation initiation is a highly regulated step of gene expression, and eIF4E is an important switch in controlling translation.[40,41] As eIF4E is the least abundant initiation factor, recruitment of eIF4E to the translation initiation complex (eIF4F) is a rate-limiting prerequisite for binding to the ribosome.[40,41] When Z binds to eIF4E it inhibits protein synthesis by preventing the formation of the eIF4F. Experimental structures of eIF4E binding to its natural target eIF4G[42] (a component of the eIF4F complex) and also to a minimal regulatory peptide E4-BP1 peptide[43] both show an α-helix recognition site [Fig. 7(a)] where the motif (**Y**DRXFLϕ) is the key interaction, ϕ is a conserved hydrophobic residue, and the most important residues are Y and L.[44] The RING domain of PML and Z is known to bind to eIF4E, but neither of these proteins contains the (**Y**DRXFLϕ) recognition motif,[45] suggesting that a yet unidentified binding interaction occurs in the eIF4E:Z complex. Therefore, we preformed protein–protein docking of Z to eIF4E in an effort to identify a binding mechanism consistent with experimental observations.

Protein–protein docking was performed using ZDOCK for all members of the $\boldsymbol{R}^{\text{Full}}$ and $\boldsymbol{H}^{\text{Full}}$ ensembles to the crystal structure of human eIF4E. The top-ranked complex of Z binding to eIF4E exhibited complementary interactions of the α-helix from Z into the eIF4E α-helix binding site [Fig. 7(b)], this will be referred to as the best static structure. Previous experiments have shown that mutations of either the first zinc center of Z (res:

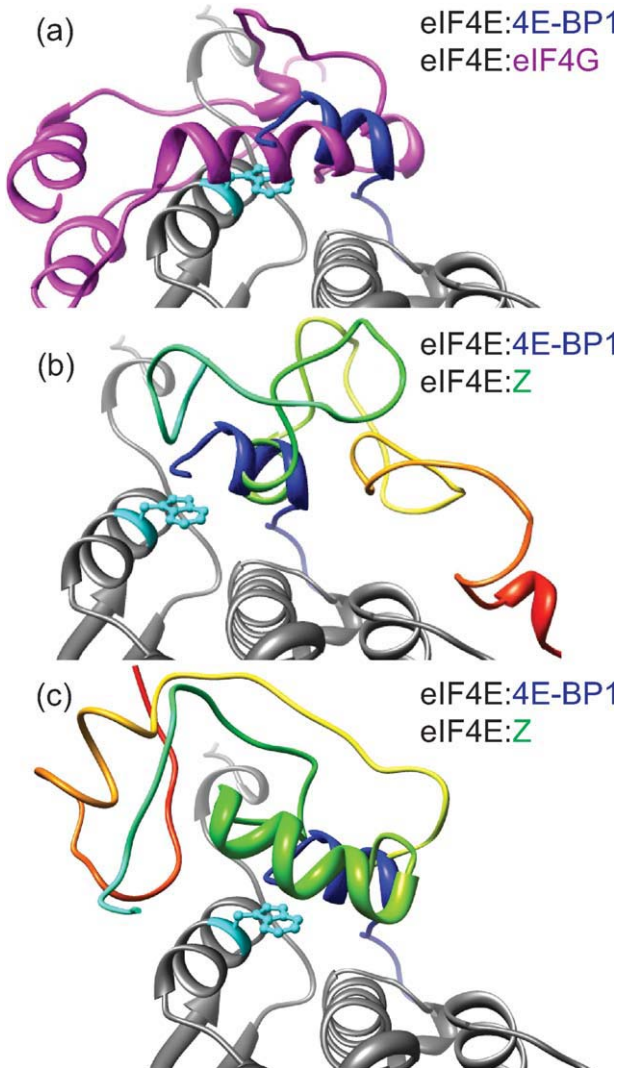

Figure 7

Protein–protein interactions between eIF4E and three binding partners, known native interactions are shown in (**a**), predicted interactions shown in (**b**) and (**c**). (a) NMR structure (1RF8) of eIF4E (gray) and a fragment of eIF4G (magenta), which is the native protein–protein interaction formed in the heterotrimeric eIF4F initiation complex. Shown in blue is the crystal structure of the fragment 4E-BP1 (1WKW) bound to eIF4E. (b, c) Predicted structures of Z bound to eIF4E from protein–protein docking. 4E-BP1 (blue) is superimposed with the helix of Z for comparison. The chain of Z is colored rainbow from blue to red to show the orientation of the C-terminus, but the N-terminal residues 1–30 are removed for clarity, as they do not interact significantly. The best static structure of Z is shown in (b) and the best dynamics structure in (c).

C31, C34, C50, and C53) or Trp73 of eIF4E also disrupted binding of Z with eIF4E.[17] Two of the top-ranked models from $\boldsymbol{R}^{\text{Full}}$ account for this observation [top ranked, Fig. 8(a) and fifth-ranked poses, Fig. 8(b)] by exhibiting a short distance between the first zinc and Trp73 of eIF4E, whereas no top-ranked models from $\boldsymbol{H}^{\text{Full}}$ agree with this observation. Both of these predicted binding modes for Z would also prevent binding of eIF4E to eIF4G and thus inhibit protein synthesis.

To further evaluate predicted binding poses, we ran five 1 ns unconstrained MD simulations for each of the top five complexes from $\boldsymbol{R}^{\text{Full}}$ and $\boldsymbol{H}^{\text{Full}}$ to check their stability. Ensemble averaged properties over the five duplicate trajectories of each complex were computed and are presented in Table III. Four of the top five poses from $\boldsymbol{R}^{\text{Full}}$ showed stable poses (<5 Å C^{α} rmsd over the core), which is in contrast to the top binding poses from $\boldsymbol{H}^{\text{Full}}$ where only one of the top five showed stability. The most stable complex was from $\boldsymbol{H}^{\text{Full}}$ (third ranked), but it did not have the most favorable interaction energy.

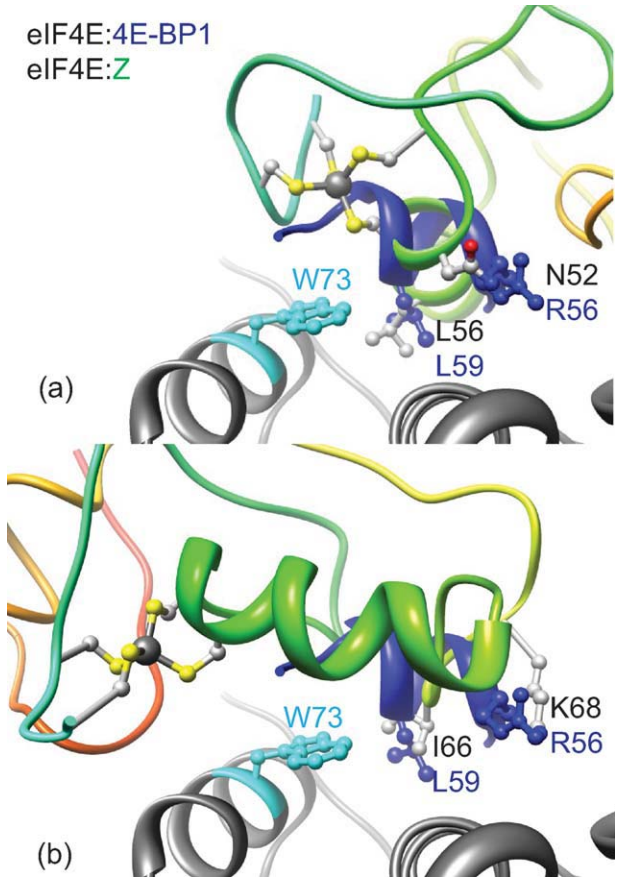

Figure 8

Detailed interactions of the protein–protein interface of eIF4E (gray) and Z (rainbow) from the best static structure complex (**a**) and best dynamics structure complex (**b**) predicted from protein–protein docking. In both of these binding modes, the first zinc center binds in close proximity to W73 (cyan) in the eIF4E structure. In both binding modes, the predicted complex of Z is shown compared to the bound structure of 4E-BP1 (blue). In the best static structure complex (a), residues from the α-helix of Z isosterically replace several important residues in the α-helical recognition motif. R56 and L59 are shown for 4E-BP1, along with the residues N52 and L59 and from Z that replace these interactions. In the best dynamics structure complex (b), Z-protein residues I66 and K68 on a loop replace these interactions. The NZ atom from the K68 side chain of Z is in very close proximity to the guanidino group of R56 of 4E-BP1.

Table III
Average Energies and rmsd Calculated for the Top Five Binding Poses from Both R^{Full} and H^{Full}

	R^{Full}					H^{Full}				
Static Rank	1	2	3	4	5	1	2	3	4	5
rmsd	4.88	4.69	4.16	9.19	4.24	6.32	7.52	3.59	12.4	10.5
ΔENER	−37.3	−56.3	−43.9	−48.5	−72.2	−54.4	−37.5	−51.8	−27.8	−32.3
ΔVDW	−43.1	−57.5	−75.5	−61.0	−83.0	−76.6	−18.1	−85.3	−28.1	−41.1
ΔELEC	−236	−299	−177	−190	−110	−143	−470	36	−202	−139
ΔGB	251	310	219	213	134	177	457	8	209	155

The rmsd was calculated by aligning the trajectories to their minimized complex and measured over the C^{α} atoms of the core (units are Å). The interaction energies listed are the total (ΔENER), the van der Waals component (ΔVDW), the electrostatic component (ΔELEC), and the electrostatic desolvation component (ΔGB) (units are kcal/mol). For each of these top binding poses, five 1 ns trajectories were run, and the averages were calculated over the last 500 ps over all five trajectories.

Additionally, this structure did not cover Trp73 of eIF4E, and there was a large separation (17.3 Å) between the first zinc center of Z and Trp73 of eIF4E; therefore, we do not consider this complex to display a likely binding mechanism. We find that the fifth-ranked pose from $\boldsymbol{R}^{\text{Full}}$ had the most favorable interaction energy and was stable during the MD simulation, and the first zinc center was proximal to Trp73 of eIF4E (9.3 Å). Based upon the proximity of the first zinc center to Trp73 of eIF4E and the stability of the trajectories, we believe that our top-ranked and fifth-ranked poses from $\boldsymbol{R}^{\text{Full}}$ are both reasonable, but that the fifth-ranked pose should be more favorable. The structure at the end of the most favorable of the five trajectories for the fifth ranked posed will be referred to the best dynamics structure. The best static structure was the fifth most populated (w_j = 7%) member of $\boldsymbol{R}^{\text{Full}}$, and the best dynamics structure was the 10th most populated (w_j = 3%) member of $\boldsymbol{R}^{\text{Full}}$. The interactions of both of these binding modes are detailed in Figure 8, showing an overly with the binding peptide 4E-BP1 and detailing the most important residue replacements.

Protein–protein interface contacts properties were calculated using a 5.25 Å cutoff and showed similar agreement with the conclusions drawn from the ensemble-averaged properties. The best dynamics structure had 3112 total contacts compared with 2364 for the crystal structure of the eIF4E:4E-BP1 complex, which agrees with the expectation that Z should bind more favorably than the 4E-BP1 regulatory peptide.[17,46] Our best static structure only exhibited 2346 total contacts, and the number of contacts decreased significantly after dynamics; in contrast, the best dynamics structure number of contacts increased during its trajectory.

A comparison of protein–protein contacts of the best static [Figs. 7(b) and 8(a)] and best dynamics structures [Figs. 7(c) and 8(b)] are shown in Table IV. We identified six residues in the Z-protein that contributed to the interface of both binding modes: residues L56, L57, and V60 of the α-helical region, and residues I92, R93, and P94 of the C-terminal region. This finding is of particular interest because these two separate sequence regions in the Z-protein are in close contact with each other in the most populated member of $\boldsymbol{R}^{\text{Full}}$. In the most populated member of $\boldsymbol{R}^{\text{Full}}$, the hydrophobic side chain of I92 on the C-terminus forms a strong side-chain contact with L56 on the α-helix, which buries the α-helical recognition motif. In this most populated conformation, the distance between the C^{β} atoms of L56 and I92 is 4.8 Å, and this distance is 20.8 and 12.9 Å for the best static and dynamics structures, respectively (see Fig. 9). In both binding modes, residue I92 from the flexible C-terminal arm moves away from the helix and forms hydrophobic contacts with eIF4E, leaving the α-helix (including L56) exposed to form strong hydrophobic contacts with eIF4E. Despite the fact that two common regions of the Z-protein contribute to the binding interface of both the best static and dynamics structures, there are also many differences in contacts between the two structures. In Table V, several mutations to the sequence of Z are proposed to discriminate between the two binding mechanism experimentally.

Table IV
Interface Contacts in the Predicted Complex of Z Binding to eIF4E

Best static structure (no dynamics)		Best dynamics structure (fifth ranked static pose)	
Z	eIF4E	Z	eIF4E
K32	E32	H25	H78
S33	W73	P28	N77
N52	E132	N46	H37
T55	R186	Y48	E70
L56	**I138**	L49	W73
L57	**H37**	C53	N77
V60	**P38**	**L56**	**Y76**
K74	D143	**L57**	**W73**
K76	E185	**V60**	**R128**
I92	**P190**	P65	W73
P93	**E185**	I66	V69
P94	**E185**	K68	E132
		I92	**H78**
		R93	**D71**
		P94	**E70**

Some example contacts between side-chain residues (heavy atom distance < 5.25 Å) are shown, which establish the overall orientation of the two binding poses. Residues of Z common to both binding poses are shown in bold.

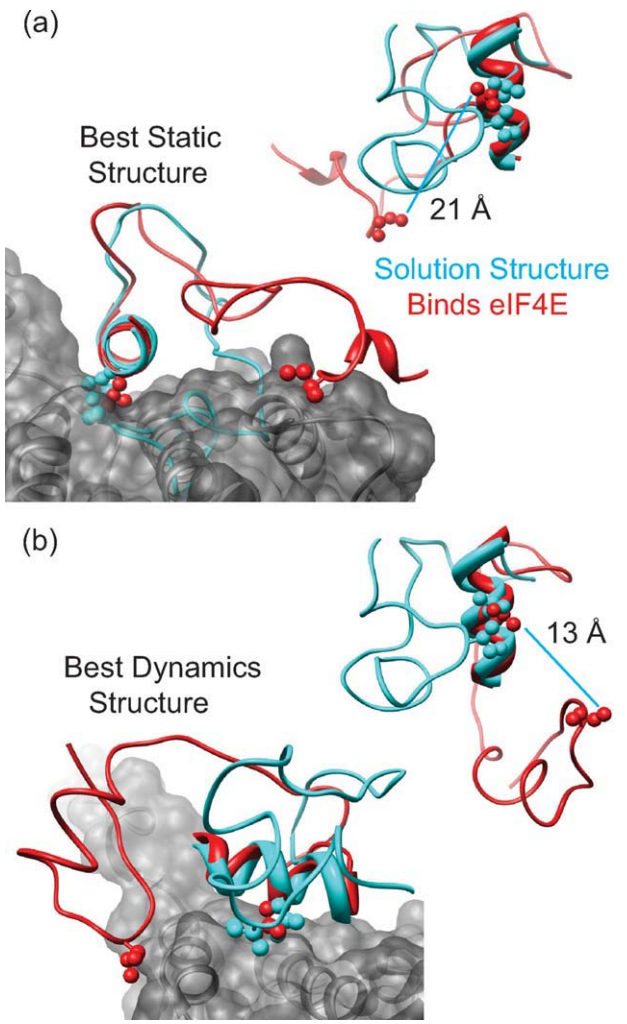

Figure 9

Comparison of predicted binding structures of Z (to eIF4E) with the most populated conformation of Z from R^{Full} (predicted solution structure). Residues L56 on the α-helix and I92 on the flexible C-terminal arm both contribute contacts to the eIF4E-binding surface. The best static structure (a) and the best dynamics structure (b) of Z in the predicted complex (red) are superimposed with the most populated solution conformation (cyan). In the solution conformation, residue I92 forms a strong hydrophobic contact with L56. In both binding conformations of Z, the flexible C-terminal arm swings out, allowing both L56 and I92 to form strong hydrophobic contacts with the surface of eIF4E (gray).

One of the key differences between various conformations of the Z-protein is the contacts formed between the C-terminal region and the core of the protein. The members of $\boldsymbol{R}^{Full}$ exhibit a variety of conformations for the C-terminus, but these different conformations are more structured and discrete (more lower order side-chain contacts), compared with the N-terminus, which behaves more as a flexible linker. The conformations of residues 68–80 are partially ordered by common hydrophobic contacts between residues L71 and L75 and the α-helical residue L58. Several of the most populated members of $\boldsymbol{R}^{Full}$ have a contact between L75 and L58, including the two most populated members and the best static binding pose to eIF4E. In the best static binding structure, the partially folded conformation of residues 68–85 is extremely important for presenting a flat interface that allows the helix to bind. Contacts that contributed to the interface from the region of res: 68–85 are a result of a partially folded and well-ordered conformation of this region, which is stabilized by several important intraprotein contacts in Z. In the best dynamics binding structure, the region of residues 74–76 formed a β-sheet structure with three well-ordered β-sheet hydrogen bonds over residues 41–43 (K74@O–E43@H, R76@H–L41@O, R76@O–L41@H). A salt bridge between R76 and E43 also stabilized this structure. In addition to the formation of this short β-sheet, this alternatively folded conformation of Z was stabilized by partially disordered helical structure over residues 80–90. In duplicate MD simulation trajectories of the best dynamics structure, this helical structure was only partially stable, likely due to the fact that it is a proline-rich region. This sequence region res: 80–90 (APTAPPTAAD) exhibits a diverse amount of partially ordered conformations with many low contact order side-chain interactions. The conformation that exhibits the structure of the PTAP motif that would allow binding to Tsg101 has broken almost all contacts between the C-terminus and the core and is likely to be disordered (see Fig. 6).

CONCLUSIONS

In this work, we have refined homology models and constructed ensembles of configurations that agree well with experimental data (chemical shifts and secondary structure) over segments and over the entire Lassa Z protein. The ensembles produce better agreement between the experimental chemical shifts than any of the individual model structures. We have identified a consistent topology and tertiary structure for the most structured region of the protein (res: 31–67) common within the

Table V

Classification of Z Protein Residues That Form Interface Contacts

Common Contacts	Best static structure		Best dynamics structure	
	Unique contacts	Mutation	Unique contacts	Mutations
L56	K32	K32D	N46	N46A
L57	S32	S32G	Y48	Y48A
V60	N52	N52A	L49	L49R
I92	K74	K74E	I66	I66A
R93	K76	K76E	K68	K68A
P94			R73	R73E

Mutations are proposed that would specifically abrogate binding in each binding mode to experimentally discriminate between the best static and dynamics binding pose.

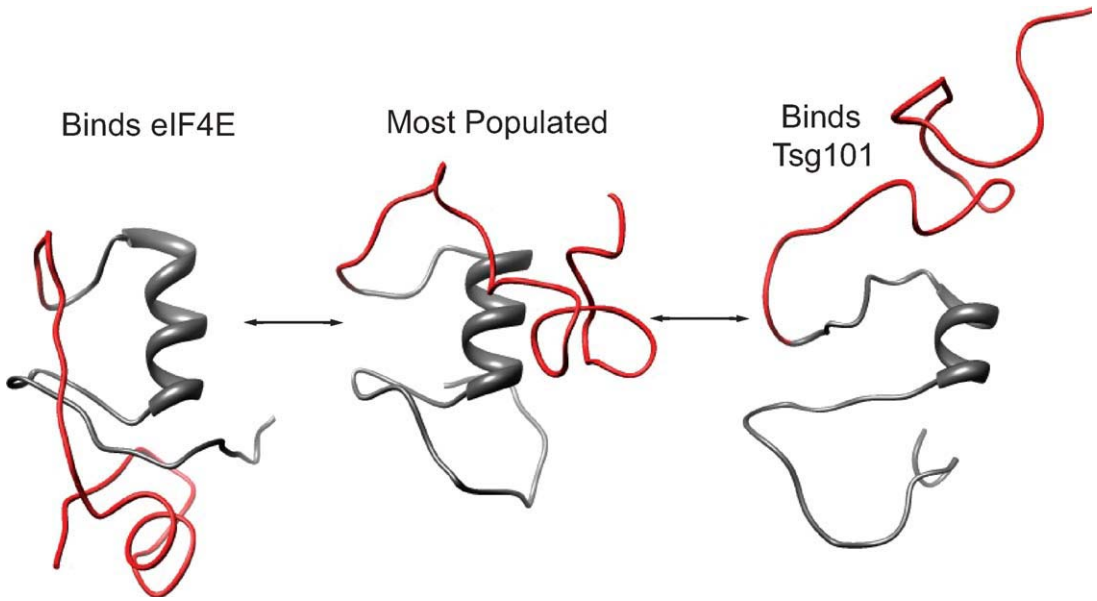

Figure 10

Three conformational states of the C-terminus: (left) conformation predicted to bind eIF4E, (center) the most populated conformation in solution, and (right) the conformation predicted to bind Tsg101.

various ensembles. Our results indicate that both the C-terminus and N-terminus are very flexible regions capable of adopting different configurations. The identification of structured and more flexible regions is consistent with experimental observations.

Our ensemble from replica exchange refined structures, $\boldsymbol{R}^{\mathrm{Full}}$, produces better agreement with the experimental chemical shifts than an ensemble of unrefined homology models $\boldsymbol{H}^{\mathrm{Full}}$. Additionally, $\boldsymbol{R}^{\mathrm{Full}}$ displays several interesting configurations not observed in the $\boldsymbol{H}^{\mathrm{Full}}$. Two structures with low-population weights in $\boldsymbol{R}^{\mathrm{Full}}$ were identified to have favorable binding with translation initiation factor eIF4E and both showed stability over a 1 ns simulation. Both structures account for the observation that Z binds at the dorsal surface of eIF4E with the first zinc center close to Trp73 of eIF4E. Common and different interface contacts between the two binding modes were identified, and mutations to Z were proposed to discriminate between the two mechanisms. The second most populated member of $\boldsymbol{R}^{\mathrm{Full}}$ displayed a configuration that we predict to be capable of binding budding factor Tsg101. This structure's extended configuration of the PTAP late domain and the lack of core-PTAP interactions present a configuration that closely matches those of the HIV-1 Gag:Tsg101 crystal structure. The most populated member of $\boldsymbol{R}^{\mathrm{Full}}$ displays a more compact configuration, which would not bind favorably to either eIF4E or Tsg101. This most populated member may represent the dominant configuration of Z when it is free in solution, absent of any binding partners. We conclude from this work that the flexible C-terminal arm of the Z-protein mediates interaction with multiple binding partners. The most populated conformation of $\boldsymbol{R}^{\mathrm{Full}}$ in solution would prevent binding to either Tsg101 or to eIF4E, whereas the second most populated conformation may bind to Tsg101, and a less populated conformation may bind to eIF4E (see Fig. 10).

MATERIALS AND METHODS

Homology modeling

Homology modeling was used to generate initial model structures of the target protein. Template structures were identified with PSI-BLAST[47]; 11 templates displayed sequence similarity to the target (25–35% identity) and also contained zinc-binding domains (pdbid: 1CHC, 1FBV, 1IYM, 2CKL:A, 2CKL:B, 2CSY, 2EA6, 2ECN, 2ECT, 2JMD, and 2YSL). The software package MODELLER[25,48] was used to build models from the templates. This package allows ligands to be built into the model structure, and therefore models were generated, which incorporated the zinc ions. For each template, five models were generated. Additionally, the sequence was submitted to the I-TASSER web server,[23,24,49] which also returned five models. The zinc ions were incorporated into the I-TASSER models by aligning these models with a MODELLER model and copying the zinc coordinates. Of the 11 templates used with MODELLER, only four (1FBV, 2CSY, 2ECN, and

2CKL:B) returned models with reasonable zinc-binding geometry. These 20 models, plus the five models returned from I-TASSER, were subjected to a structure refinement protocol.

Energy evaluations and replica exchange MD with CHARMM

Before any molecular mechanics were performed, the zinc-binding cysteines (res: 31, 34, 44, 50, 53, 64, and 67) were modified by deleting the sulfur-bound hydrogen and all associated bonds, angles, and dihedrals. The force field parameters of the zinc-binding cysteines were made consistent with the model described by Bredenberg and Nilsson.[50] The CHARMM simulation package,[51,52] with the CHARMM27/CMAP all-atom force field,[53–55] was used for all minimization and dynamics calculations in this work. A total of 2.4 Å separations between zinc and the zinc-coordinating atoms (sulfur for cysteine, δ-nitrogen for histidine) were maintained with 100 kcal/(mol Å^2) harmonic restraints. The hydrogen-bond distances were equilibrated with 1000 steps of steepest descent (SD) minimization in vacuum with restraints on all non-hydrogen atoms; successive rounds of minimization with 300, 200, and 100 kcal/(mol Å^2) spring constants were used. The same procedure was repeated using the GBMV implicit solvent model,[56–58] with 100 SD steps at each restraint level. A 25 kcal/(mol Å^2) harmonic restraint was placed on all backbone atoms, and 100 SD steps were performed followed by 100 steps of adopted basis Newton-Raphson.

The 25 minimized structures were each submitted to a REX protocol[27,28] carried out using the aarex.pl script, available from the Multiscale Modeling Tools in Structural Biology (MMTSB Tool Set).[26] Sixteen temperature windows ranging from 300 to 600 K were used. Each replica was run for 500 fs between exchange attempts; 1000 REX cycles were done for each model. The dynamics of each replica was conducted using GBMV implicit solvent, the SHAKE algorithm[59] to fix hydrogen-bond lengths, 100 kcal/(mol Å^2) harmonic restraints between zinc and zinc-coordinating atoms, and a 1 fs time step. Both the van der Waals and electrostatic interactions were cutoff at 16 Å.

The structures from the last 101 REX cycles in the lowest temperature window (300 K) for all models were gathered for clustering (2525 total structures). A hierarchal clustering method was performed using the cluster.pl program available in the MMTSB Tool Set.[26] The clusters were delineated based upon rmsd values of the C$^\alpha$ and C$^\beta$ atoms. Further clustering was attempted on any cluster with at least 50 members; clusters were subclustered into a maximum of 100 clusters. The energy was evaluated for all members of each cluster using the same parameters as the REX dynamics runs, and the lowest energy structure in each cluster was identified. Chemical shifts of the structures were computed using the SPARTA program, which computes chemical shifts for backbone N, H^N, H$^\alpha$, C$^\alpha$, C$^\beta$, and C′ atoms.[34] Helical segments were identified using DSSP[60] and the UCSF Chimera program.[61]

Construction of minimal CS-rmsd ensembles

A CS-rmsd was calculated between the predicted and experimental CS. A nonlinear fitting algorithm was used to minimize the objective function χ^2, shown in Eq. (1), subject to the constraint of Eq. (2)

$$\chi^2 = \sum_i \left(\left(\sum_j w_j \mathrm{CS}_{j,i}^{\mathrm{Comp}} \right) - \mathrm{CS}_i^{\mathrm{Exp}} \right)^2 \qquad (1)$$

$$\sum_j w_j = 1 \ , \qquad (2)$$

where w_j is the weighting factor given to each structure, the sum i is over atoms, and j over the number of structures considered in the averaging. The weighting factor can be interpreted as the population weight of a configuration (i.e., the contribution of a structure to the ensemble). To test the robustness of the fitting algorithm, we randomly chose 13 of the 25 homology model structures, assigned each an arbitrary weight [obeying Eq. (2)], and computed a chemical shifts test set from these structures and weights. We then used this chemical shifts test set in place of the experimental shifts in Eq. (1) and ran the fitting algorithm on the full set of 25 homology models. The fitting algorithm was successful in assigning nonzero weights to only the 13 structures used to construct the test set, returned the same weights as the input weights (within 10^{-3}), and gave a CS-rmsd of 2×10^{-4} ppm. It should be noted that the threshold value used throughout this article to determine nonzero weights was $w_j \geq 10^{-2}$ (1%).

Protein–protein docking

The models of Z were docked against the crystal structure of eIF4E (1WKW). The eIF4E complex[43] was prepared for docking by removing the coordinates of all crystallographic waters as well as the bound regulatory peptide 4E-BP1 and the 5′ mRNA cap analog m7GpppA. All members of the $\boldsymbol{R}^{\mathrm{Full}}$ and $\boldsymbol{H}^{\mathrm{Full}}$ were docked against the prepared structure of eIF4E. Fully unconstrained rigid-body docking was performed using the ZDOCK web server (http://zdock.bu.edu).[62–65] For each of the Z models in either ensemble, the top 50 binding poses out of 1000 (as ranked by ZDOCK) were then subjected to energy minimization in CHARMM using a 100 kcal/(mol Å^2) harmonic restraint applied to the internal coor-

dinates of both Z and eIF4E. First, 1000 steps of SD minimization was performed with a soft-core potential function followed by 1000 steps of SD minimization with the standard potential function. The GBMV implicit solvent model was applied, and an additional 500 steps of SD minimization was performed. The restraints on the internal coordinates were released, and interaction energies of the complexes were calculated. The total interaction energy was used to rank all of the complex structures. The interaction energy was computed by calculating the energy of the eIF4E:Z complex and then subtracting the energies of the individual unbound eIF4E and Z structures. The only contribution to the interaction energy came from the nonbonded terms (van der Waals, electrostatic, and GBMV), as the configuration of the individual proteins in the complexed and free state is identical.

ACKNOWLEDGMENTS

Molecular graphics images were produced using the UCSF Chimera[61] package from the Resource of Biocomputing, Visualization, and Informatics at the University of California, San Francisco.

REFERENCES

1. Borio L, Inglesby T, Peters CJ, Schmaljohn AL, Hughes JM, Jahrling PB, Ksiazek T, Johnson KM, Meyerhoff A, O'Toole T, Ascher MS, Bartlett J, Breman JG, Eitzen EM, Hamburg M, Hauer J, Henderson A, Johnson RT, Kwik G, Layton M, Lillibridge S, Nabel GJ, Osterholm MT, Perl TM, Russell P, Tonat K;Working Group Civilian B. Hemorrhagic fever viruses as biological weapons—medical and public health management. J Am Med Assoc 2002;287:2391–2405.
2. Gunther S, Lenz O. Lassa virus. Crit Rev Clin Lab Sci 2004;41:339–390.
3. Fulhorst CF, Bennett SG, Milazzo ML, Murray HL, Jr, Webb JP, Jr, Cajimat MN, Bradley RD. Bear Canyon virus: an arenavirus naturally associated with the California mouse (*Peromyscus californicus*). Emerg Infect Dis 2002;8:717–721.
4. Fulhorst CF, Charrel RN, Weaver SC, Ksiazek TG, Bradley RD, Milazzo ML, Tesh RB, Bowen MD. Geographic distribution and genetic diversity of Whitewater Arroyo virus in the southwestern United States. Emerg Infect Dis 2001;7:403–407.
5. Calisher CH, Nabity S, Root JJ, Fulhorst CF, Beaty BJ. Transmission of an arenavirus in white-throated woodrats (*Neotoma albigula*), southeastern Colorado, 1995-1999. Emerg Infect Dis 2001;7:397–402.
6. Armstrong LR, Dembry LM, Rainey PM, Russi MB, Khan AS, Fischer SH, Edberg SC, Ksiazek TG, Rollin PE, Peters CJ. Management of a Sabia virus-infected patients in a US hospital. Infect Control Hosp Epidemiol 1999;20:176–182.
7. Bray M. Highly pathogenic RNA viral infections: challenges for antiviral research. Antiviral Res 2008;78:1–8.
8. Bowden TA, Crispin M, Graham SC, Harvey DJ, Grimes JM, Jones EY, Stuart DI. Unusual molecular architecture of the Machupo virus attachment glycoprotein. J Virol 2009;83:8259–8265.
9. Eichler R, Strecker T, Kolesnikova L, ter Meulen J, Weissenhorn W, Becker S, Klenk HD, Garten W, Lenz O. Characterization of the Lassa virus matrix protein Z: electron microscopic study of virus-like particles and interaction with the nucleoprotein (NP). Virus Res 2004;100:249–255.
10. Strecker T, Maisa A, Daffis S, Eichler R, Lenz O, Garten W. The role of myristoylation in the membrane association of the Lassa virus matrix protein Z. Virol J 2006;3:93.
11. Perez M, Craven RC, de la Torre JC. The small RING finger protein Z drives arenavirus budding: implications for antiviral strategies. Proc Natl Acad Sci USA 2003;100:12978–12983.
12. Garrus JE, von Schwedler UK, Pornillos OW, Morham SG, Zavitz KH, Wang HE, Wettstein DA, Stray KM, Cote M, Rich RL, Myszka DG, Sundquist WI. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. Cell 2001;107:55–65.
13. Timmins J, Schoehn G, Ricard-Blum S, Scianimanico S, Vernet T, Ruigrok RWH, Weissenhorn W. Ebola virus matrix protein VP40 interaction with human cellular factors Tsg101 and Nedd4. J Mol Biol 2003;326:493–502.
14. Borden KLB, Dwyer EJC, Salvato MS. An arenavirus RING (zinc-binding) protein binds the oncoprotein promyelocyte leukemia protein (PML) and relocates PML nuclear bodies to the cytoplasm. J Virol 1998;72:758–766.
15. Topcu Z, Mack DL, Hromas RA, Borden KL. The promyelocytic leukemia protein PML interacts with the proline-rich homeodomain protein PRH: a RING may link hematopoiesis and growth control. Oncogene 1999;18:7091–7100.
16. Borden KL, Campbelldwyer EJ, Carlile GW, Djavani M, Salvato MS. Two RING finger proteins, the oncoprotein PML and the arenavirus Z protein, colocalize with the nuclear fraction of the ribosomal P proteins. J Virol 1998;72:3819–3826.
17. Kentsis A, Dwyer EC, Perez JM, Sharma M, Chen A, Pan ZQ, Borden KLB. The RING domains of the promyelocytic leukemia protein PML and the arenaviral protein Z repress translation by directly inhibiting translation initiation factor eIF4E. J Mol Biol 2001;312:609–623.
18. Volpon L, Osborne MJ, Borden KL. NMR assignment of the arenaviral protein Z from Lassa fever virus. Biomol NMR Assign 2008;2:81–84.
19. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. Protein structure determination from NMR chemical shifts. Proc Natl Acad Sci USA 2007;104:9615–9620.
20. Bradley P, Misura KMS, Baker D. Toward high-resolution de novo structure prediction for small proteins. Science 2005;309:1868–1871.
21. Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu GH, Eletsky A, Wu YB, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A. Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci USA 2008;105:4685–4690.
22. Shen Y, Vernon R, Baker D, Bax A. De novo protein structure generation from incomplete chemical shift assignments. J Biomol NMR 2009;43:63–78.
23. Zhang Y. Template-based modeling and free modeling by I-TASSER in CASP7. Proteins 2007;69 (Suppl 8):108–117.
24. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 2008;9:40.
25. Sali A, Blundell TL. Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
26. Feig M, Karanicolas J, Brooks CL, III. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graph Model 2004;22:377–395.
27. Chen JH, Im W, Brooks CL, III. Refinement of NMR structures using implicit solvent and advanced sampling techniques. J Am Chem Soc 2004;126:16038–16047.
28. Chen JH, Won HS, Im WP, Dyson HJ, Brooks CL, III. Generation of native-like protein structures from limited NMR data, modern force fields and advanced conformational sampling. J Biomol NMR 2005;31:59–64.
29. Xu XP, Case DA. Automated prediction of N-15. C-13(alpha), C-13(beta) and C-13' chemical shifts in proteins using a density functional database. J Biomol NMR 2001;21:321–333.

30. Xu XP, Case DA. Probing multiple effects on N-15. C-13 alpha C-13 beta, and C-13' chemical shifts in peptides using density functional theory. Biopolymers 2002;65:408–423.
31. Neal S, Nip AM, Zhang HY, Wishart DS. Rapid and accurate calculation of protein H-1. C-13 and N-15 chemical shifts. J Biomol NMR 2003;26:215–240.
32. Meiler J. PROSHIFT: protein chemical shift prediction using artificial neural networks. J Biomol NMR 2003;26:25–37.
33. Vila JA, Arnautova YA, Martin OA, Scheraga HA. Quantum-mechanics-derived C-13(alpha) chemical shift server (CheShift) for protein structure validation. Proc Natl Acad Sci USA 2009;106:16972–16977.
34. Shen Y, Bax A. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J Biomol NMR 2007;38:289–302.
35. Choy WY, Forman-Kay JD. Calculation of ensembles of structures representing the unfolded state of an SH3 domain. J Mol Biol 2001;308:1011–1032.
36. Krzeminski M, Fuentes G, Boelens R, Bonvin AMJJ. MINOES: a new approach to select a representative ensemble of structures in NMR studies of [partially] unfolded states. Application to Delta 25-PYP. Proteins: Struct Funct Bioinformatics 2009;74:895–904.
37. Strecker T, Eichler R, ter Meulen J, Weissenhorn W, Klenk HD, Garten W, Lenz O. Lassa virus Z protein is a matrix protein sufficient for the release of virus-like particles. J Virol 2003;77:10700–10705.
38. Banci L, Bertini L, Ciofi-Baffoni S, Finney LA, Outten CE, O'Halloran TV. A new zinc-protein coordination site in intracellular metal trafficking: solution structure of the Apo and Zn(II) forms of ZntA(46-118). J Mol Biol 2002;323:883–897.
39. Pornillos O, Alam SL, Davis DR, Sundquist WI. Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein. Nat Struct Biol 2002;9:812–817.
40. Sachs AB, Sarnow P, Hentze MW. Starting at the beginning, middle, and end: translation initiation in eukaryotes. Cell 1997;89:831–838.
41. Raught B, Gingras AC. eIF4E activity is regulated at multiple levels. Int J Biochem Cell B 1999;31:43–57.
42. Gross JD, Moerke NJ, von der Haar T, Lugovskoy AA, Sachs AB, McCarthy JE, Wagner G. Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell 2003;115:739–750.
43. Tomoo K, Matsushita Y, Fujisaki H, Abiko F, Shen X, Taniguchi T, Miyagawa H, Kitamura K, Miura K, Ishida T. Structural basis for mRNA cap-binding regulation of eukaryotic initiation factor 4E by 4E-binding protein, studied by spectroscopic. X-ray crystal structural, and molecular dynamics simulation methods. Biochim Biophys Acta 2005;1753:191–208.
44. Poulin F, Gingras AC, Olsen H, Chevalier S, Sonenberg N. 4E-BP3, a new member of the eukaryotic initiation factor 4E-binding protein family. J Biol Chem 1998;273:14002–14007.
45. Cohen N, Sharma M, Kentsis A, Perez JM, Strudwick S, Borden KLB. PML RING suppresses oncogenic transformation by reducing the affinity of eIF4E for mRNA. EMBO J 2001;20:4547–4559.
46. Niedzwiecka A, Marcotrigiano J, Stepinski J, Jankowska-Anyszka M, Wyslouch-Cieszynska A, Dadlez M, Gingras AC, Mak P, Darzynkiewicz E, Sonenberg N, Burley SK, Stolarski R. Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5′ cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. J Mol Biol 2002;319:615–635.
47. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
48. Fiser A, Do RKG, Sali A. Modeling of loops in protein structures. Protein Sci 2000;9:1753–1773.
49. Wu ST, Skolnick J, Zhang Y. Ab initio modeling of small proteins by iterative TASSER simulations. BMC Biol 2007;5:17.
50. Bredenberg J, Nilsson L. Modeling zinc sulfhydryl bonds in zinc fingers. Int J Quantum Chem 2001;83:230–244.
51. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
52. Brooks BR, Brooks CL, III, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. J Comput Chem 2009;30:1545–1614.
53. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
54. Mackerell AD, Feig M, Brooks CL, III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 2004;25:1400–1415.
55. MacKerell AD, Feig M, Brooks CL, III. Improved treatment of the protein backbone in empirical force fields. J Am Chem Soc 2004;126:698–699.
56. Lee MS, Feig M, Salsbury FR, Brooks CL, III. New analytic approximation to the standard molecular volume definition and its application to generalized born calculations. J Comput Chem 2003;24:1348–1356.
57. Lee MS, Salsbury FR, Brooks CL, III. Novel generalized Born methods. J Chem Phys 2002;116:10606–10614.
58. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL, III. Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 2004;25:265–284.
59. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-integration of Cartesian equations of motion of a system with constraints—molecular-dynamics of N-alkanes. J Comput Phys 1977;23:327–341.
60. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
61. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–1612.
62. Chen R, Li L, Weng ZP. ZDOCK: an initial-stage protein-docking algorithm. Proteins 2003;52:80–87.
63. Chen R, Mintseris J, Janin J, Weng ZP. A protein-protein docking benchmark. Proteins 2003;52:88–91.
64. Chen R, Tong WW, Mintseris J, Li L, Weng ZP. ZDOCK predictions for the CAPRI challenge. Proteins 2003;52:68–73.
65. Chen R, Weng ZP. A novel shape complementarity scoring function for protein-protein docking. Proteins 2003;51:397–408.