Disordered Flanks Prevent Peptide Aggregation

Sanne Abeln*, Daan Frenkel

FOM Institute for Atomic and Molecular Physics, Amsterdam, The Netherlands

Abstract

Natively unstructured or disordered regions appear to be abundant in eukaryotic proteins. Many such regions have been found alongside small linear binding motifs. We report a Monte Carlo study that aims to elucidate the role of disordered regions adjacent to such binding motifs. The coarse-grained simulations show that small hydrophobic peptides without disordered flanks tend to aggregate under conditions where peptides embedded in unstructured peptide sequences are stable as monomers or as part of small micelle-like clusters. Surprisingly, the binding free energy of the motif is barely decreased by the presence of disordered flanking regions, although it is sensitive to the loss of entropy of the motif itself upon binding. This latter effect allows for reversible binding of the signalling motif to the substrate. The work provides insights into a mechanism that prevents the aggregation of signalling peptides, distinct from the general mechanism of protein folding, and provides a testable hypothesis to explain the abundance of disordered regions in proteins.

Introduction

The biological function of many proteins is determined by their native, three-dimensional structure and unfolded (or incorrectly folded) copies of such proteins tend to be inactive, if not outright dangerous.

However, many proteins contain large regions (>30 amino acids) that are disordered in their natural physico-chemical environment [1–4]; some proteins are even entirely disordered [5,6]. As more peptide sequences are being studied, it is becoming increasingly clear that natively-disordered sequences are far more common than previously thought. Disordered sequences have been found on a large number of eukaryotic genes (>30%) [2,5,7,8]. Moreover, the number of genes on a genome with disordered regions appears to increase with the complexity of the species [2,5,7,8].

Despite a lack of stable structure in the native form of the protein, disorder is strongly associated with specific cellular functions, most significantly with cell signalling and regulatory processes [9–14]. Several suggestions have been made about the possible benefits of disordered regions in a protein: they could be more malleable, have a large binding surface, bind to diverse ligands, bind with high specificity and make the binding process reversible [1,12,15,16]. Indeed, there exist numerous examples of natively disordered proteins that form a more defined structure upon binding to a ligand [17], implying that the protein loses conformational entropy on binding.

Disordered regions (peptide sequences that are generally unfolded) and natively unstructured binding regions (sequences that only take a specific structure upon binding) have some general features. Disordered regions contain fewer hydrophobic, more hydrophilic, more charged amino acids and more repeats in their sequence as compared to natively structured proteins [6].

On the other hand interfacial regions between a natively unstructured binding region and a rigid protein contain relatively more hydrophobic and fewer charged contacts, as compared to rigid-rigid interfaces [18]. In general, only a small (hydrophobic) motif of the disordered region is involved in the actual binding and this binding motif remains in an extended configuration even upon binding and 'folding' [19–21]. As a consequence, the exposed binding area per residue is relatively large [15,18] (see Figure 1).

Recent studies have revealed that many small (linear) binding motifs are surrounded by disordered regions [22,23]. A typical linear binding motif contains some 6 residues and is surrounded by approximately 20 residues that are natively unstructured [23]. The binding motifs are typically more hydrophobic than the flanking residues. Since the binding regions are relatively small, they are unlikely to form fully folded (or specific) structures in solution when not bound to a substrate. In this study we focus on the steric effects of the disordered regions adjacent to small hydrophobic binding motifs.

As the presence of disordered regions near small binding motifs appears to be generic, it seems justified to use a generic model. The nature of the coarse-grained model allows us to simulate the specificity, steric hindrance, configurational and translational entropy of the peptide chain. Each residue of the peptide chain occupies a single point on a cubic lattice. The lattice makes efficient movements in the peptide chain possible so that many different configurations of the chain can be sampled with a Monte Carlo algorithm. Residues on neighbouring lattice points interact in a pairwise manner. Each of the 20 amino acids has a specific interaction energy with each of the other amino acids [24,25]. For example, two neighbouring hydrophobic amino acids lower the internal energy and are thus attracted to each other. The large number of possible interactions and sequences enables the design of amino acid sequences that fold into a specific structure [26,27]. Using these designed peptide sequences it is possible to describe the folding mechanism of highly specific folding [26,27] or binding [16,28]. However, due to its coarse-grained nature, the model

Author Summary

In their natural cellular environment proteins are dissolved in a concentrated aqueous solution of biomolecules. Even under such crowded conditions, proteins must not clump together or aggregate; otherwise their biological functions may be compromised, and the cell could die. Diseases such as Parkinson and Alzheimer are thought to be caused by aggregation of specific proteins. Evolutionary pressure generally ensures that proteins do not aggregate in their natural biochemical environment. A well-known mechanism to prevent aggregation is the folding of proteins, where the hydrophobic (attractive) part of the protein is buried inside the protein. Here we report a different mechanism that can prevent the aggregation of proteins. Recently, it was discovered that many proteins contain regions that are disordered (not folded) in their natural environment. We show with coarse-grained simulations that aggregation of small hydrophobic binding motifs can be prevented by embedding the motifs in disordered regions: the disordered regions of different proteins obstruct or sterically hinder the formation of aggregates. Moreover, our simulations show that the disordered regions have no adverse effect on the biological function of the binding motifs, because they do not obstruct the binding and folding of the binding motif on its specific substrate.

would be unsuited to represent the structure or binding site of a specific, naturally occurring protein.

We use this coarse-grained model to investigate how the binding free energy of a short binding motif depends upon its structural environment: we simulate binding to a substrate for a flexible binding motif, a flexible motif embedded in an unstructured chain and a rigid binding motif embedded in a rigid structure (see Figure
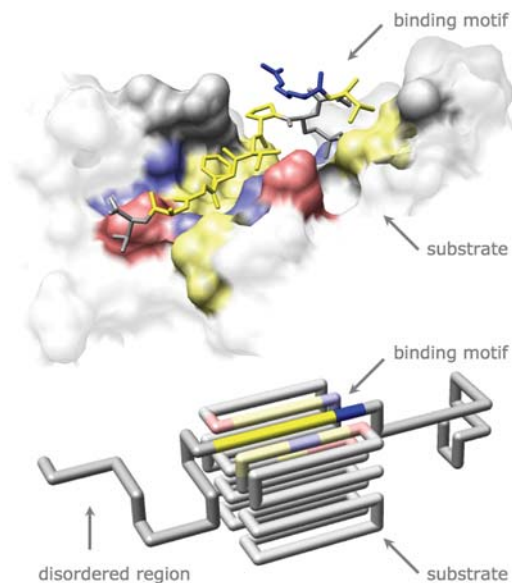
Figure 1. Linear binding motifs. Top: Example of a linear binding motif bound to its substrate. CtIP phosphopeptide is bound to BRCT repeats of BRCA1 (1Y98). Bottom: Model of a binding motif. The motif, with sequence RWWLY, is designed to bind specifically to the substrate. The yellow residues are hydrophobic, the blue negatively charged, the red positively charged and the grey hydrophilic.
doi:10.1371/journal.pcbi.1000241.g001

S1). The model of the substrate and binding region embedded in disordered flanks have been designed to contain the general features associated with disordered regions and natively unstructured binding regions, viz. an extended binding conformation, a large binding surface, hydrophobicity of the binding region and hydrophilic flanks.

We find that the binding motif embedded in a rigid structure unbinds at higher temperatures than either the flexible binding motif or the binding motif in a longer disordered region. The latter two binding free energies are very similar over the range of temperatures simulated. However, we show that even at low concentrations the (hydrophobic) binding motif aggregates with itself, and that the (hydrophilic) disordered flanks prevent such aggregation at temperatures relevant for reversible binding.

Results

Folding and Binding of Binding Motifs

To investigate how the binding free energy of a short binding motif depends upon its structural environment, a binding motif was designed to specifically bind in a groove of a rigid substrate (Figure 1). The amino acid sequence (Arg, Trp, Tr, Leu, Tyr) of this motif is predominantly hydrophobic, but contains a single charged amino acid. In our coarse-grained model, neighbouring hydrophobic residues attract each other, whereas amino acids of the same charge repel each other.

The binding of this binding motif was simulated embedded in three different structures: as a single flexible binding motif (BM), as a single flexible binding motif with disordered flanks of 15 Threonine residues on each side (BM disorder) and embedded in a rigid structure of Threonine residues (BM rigid), see Figures S1 and S2. Threonine is a hydrophilic amino acid. In our model contacts involving Threonine do not contribute to the internal energy of the configuration so that the internal energy of the binding motif bound to the substrate is the same for all three structures (see Methods).

The binding and unbinding process was simulated at different temperatures, while the concentration of the substrate and peptide are kept constant. Figure 2 shows that at low temperatures ($T<0.25$) the average degree of binding ($\langle P_b \rangle$) is high, i.e. the binding motif is nearly always bound to the substrate, and at high temperatures ($T>0.45$) the average degree of binding is low. The flexible peptides (BM and BM disorder) are unstructured in the unbound state (see Figure S2).

There is a transition between the bound and unbound state at which reversible binding is possible. This transition can also be observed by the peak in the heat capacity (C_v). Similar peaks in heat capacity are found at folding transitions of both simulated and real proteins (e.g., [29,30]). The sharpness of the heat-capacity curve also indicates that the binding motif binds with high specificity to the substrate. Binding of an aspecific motif to the substrate would result in a much broader heat-capacity peak.

In nature binding motifs typically have a signalling function, implying that the peptide should be able to bind as well as unbind in the relevant temperature range. Figure 2 shows that the binding motif binds reversibly to the substrate for approximately $0.2<T<0.3$.

Interestingly, Figure 2 shows that the disordered flanks have little effect on the binding free energy: the average amount of binding and heat capacity are similar over the entire temperature range for both flexible peptides (BM and BM disorder). Additional simulations showed that even with a much larger substrate the difference in binding free energy between the binding motif and the motif embedded in disordered flanks remains small. However,
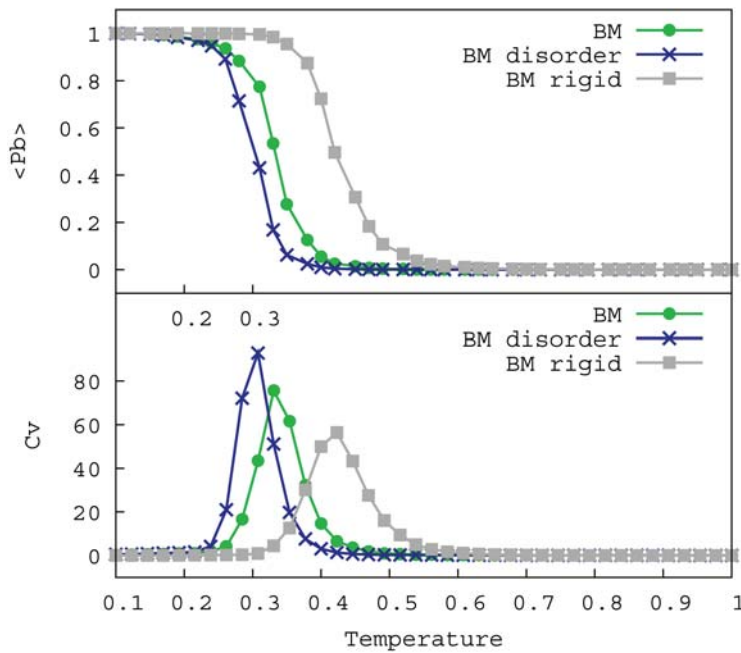
Figure 2. Reversible binding. Average amount of binding ($\langle P_b \rangle$, top) and heat capacity (C_v, bottom) as a function of temperature shown for an isolated binding motif (BM), a binding motif within disordered flanks (BM disorder) and a rigid binding motif embedded in a rigid structure (BM rigid).
doi:10.1371/journal.pcbi.1000241.g002

as previously reported [16], the flexibility of the binding motif itself lowers the difference in free energy between the bound and unbound state, since conformational entropy is lost upon binding to the substrate. Figure 2 shows that the temperature range for reversible binding of flexible peptide chains is lower than for a rigid binding motif.

Aggregation of Small Binding Peptides

Even though disordered flanks appear to contribute little to the binding free energy, the collective contribution of many such flanks may be important. We simulated 10 binding motifs without the substrate to investigate the collective behaviour of the peptides. Figure 3 shows that 10 binding motifs without flanks tend to aggregate whereas those with flanks do not at a temperature at which reversible binding is possible; the lowest free energy configuration for 10 binding motifs with flanks is as free chains or in very small clusters, whereas the binding motifs without flanks make many more external contacts.

To investigate this phenomenon for a larger number of peptide chains, we simulated aggregation behaviour of the two types of binding motifs with a Grand Canonical Monte Carlo simulation, while keeping the free binding motifs at low concentration (see Methods).

First, simulations starting from a single chain in the simulation box were performed at different temperatures. Many more external contacts form for the binding motif than for the binding motif embedded in disordered flanks (Figure 4). Moreover, the aggregates form at higher temperatures for binding motifs without disordered flanks. From these simulations we selected aggregates of different cluster sizes. Each cluster of aggregates was simulated at different temperatures to determine the transition temperature, T_s, at which the aggregate would shrink rather than grow in size (Figure 5).

Comparing Figure 2 with Figure 5 it can be observed that the binding motifs (BM) are in an aggregated state at temperatures within the reversible binding regime, whereas the binding motifs with disordered (BM disorder) are fully dissolved. Figure 2 also shows
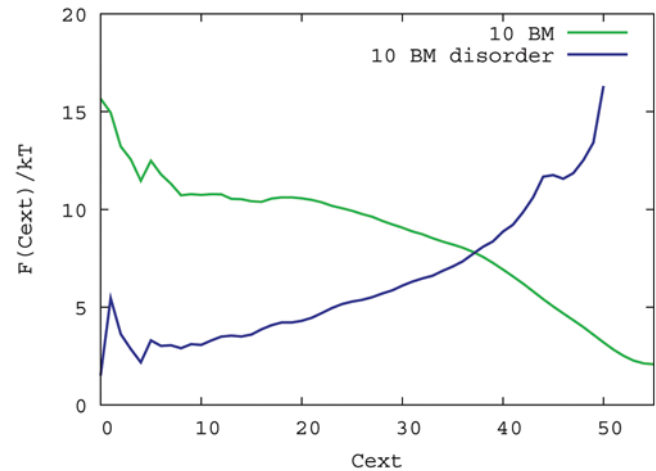

Figure 3. Aggregation free energy. Free energy as a function of external contacts at T = 0.23. The free energy is defined as $F(C_{ext}) = -k_B T \ln(P(C_{ext}))$ where $P(C_{ext})$ is the probability of a configuration with C_{ext} external contacts. The number of peptide chains was kept constant at 10. Free energies for $0 < C_{ext} < 55$ are displayed; free energies for a higher number of external contacts are dominated by finite size effects (10 peptides) effect of the system.
doi:10.1371/journal.pcbi.1000241.g003

that with increasing aggregate size the aggregates formed by binding motifs without disordered flanks become more difficult to melt, indicating that once an aggregate is formed it will be difficult to dissolve. Binding motifs embedded in disordered domains, generally form micelle-like structures that do not grow larger than approximately 12 chains (see Figure 4). Decreasing the length of the disordered flanks, down to 5 residues on each side of the binding motif, does not have a strong effect on the melting temperatures. In that case the micelles formed are somewhat larger.

The system also shows considerable hysteresis: the aggregated clusters melt at much higher temperatures than the ones at which
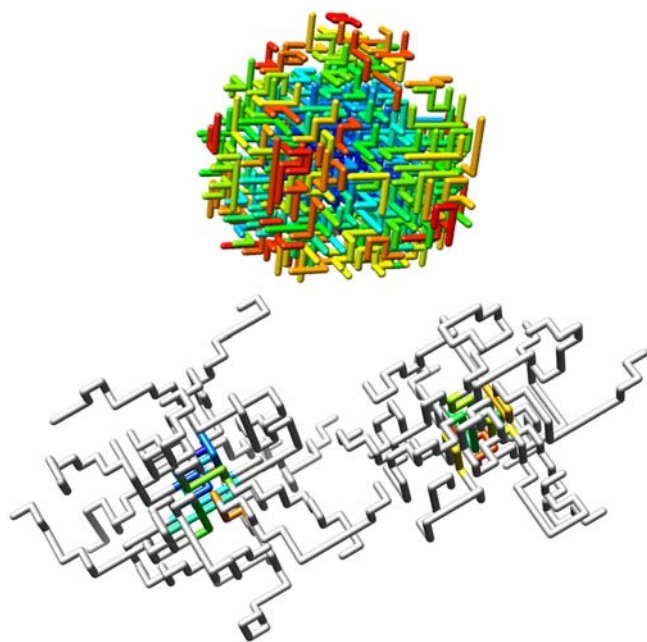
Figure 4. Aggregation of binding motifs. Top: snapshot of 301 aggregated binding motifs. Bottom: snapshot of two micelles formed by 18 binding motifs embedded in Threonine flanks (grey). The binding motifs have been given a colour ranging from blue to red according to their order of appearance in the simulation box.
doi:10.1371/journal.pcbi.1000241.g004
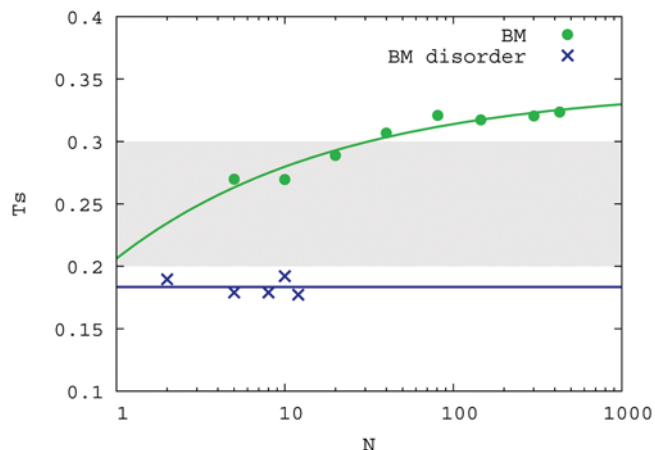
Figure 5. Melting temperatures of aggregated clusters. Cluster size (N) versus melting temperatures (T_s) for different cluster sizes. The shaded area indicates the temperature range in which reversible binding is possible for the flexible binding motifs to the substrate (see Figure 2). Stable aggregates exist in the regions below the melting curves.
doi:10.1371/journal.pcbi.1000241.g005

they formed. Again, this effect is much smaller for binding motifs embedded in disordered flanks.

Discussion

Our simulations suggest that the primary role of disordered flanks adjacent to small peptide binding motifs is to suppress aggregation in solution rather than to modify the binding strength to the substrate. This observation provides a rationale for the experimental observation that linear binding motifs are often found in disordered parts of a peptide chain [23].

In this work only a small difference in binding strength between binding motifs with and without disordered flanks is found. The model used here is based on the assumption that interactions between the disordered flanks and the substrate are of a steric nature. However our results do not preclude the possibility that the binding strength changes significantly if the disordered flanks have additional interactions with the substrate, for example through charged residues or a second binding motif. Our work focuses on the physical effect of disordered flanks that have no specific interaction with the substrate.

The isolated binding motifs described in the present paper would aggregate due to hydrophobic interactions. We suggest that such motifs, without hydrophilic flanks, are toxic. There is indeed increasing evidence that hydrophobic aggregation is correlated with toxicity for the cell [31]. Of course, the model calculations that we present here are highly simplified. The degree of hydrophobicity in real binding motifs varies, although it is typically higher than that of disordered proteins or that of the surface of globular proteins. There is, therefore, a great need for experiments to quantify the difference in aggregation behavior of signalling peptides with and without disordered flanks.

Aggregated proteins can form different structures: ordered beta sheet fibers (amyloids) or non-specific hydrophobic aggregates. Human diseases, such as Alzheimer and Parkinson disease, are mostly associated with the former. The work presented here is most closely related to the latter mechanism. Nevertheless, there is increasing evidence that the two mechanisms are connected and that hydrophobic pre-fibrillar aggregates may be causing the toxicity in amyloid forming proteins [32,33]. Insights in (the prevention of) protein hydrophobic aggregation may therefore be important for further understanding of both aggregation types.

Of course, there could be other ways to suppress hydrophobic aggregation. For instance, aggregation would be strongly inhibited if the binding motif were embedded in a rigid structure [34]. However, a flexible binding motif has the advantage that it can combine the ability to bind reversibly with high specificity: this feature is important for regulatory motifs.

As such, it would not be surprising to find that disordered flanks have evolved to suppress aggregation. There are several other biological examples of evolutionary pressure against aggregation [34]. For example: there exist very few proteins with beta-strands on the edge of protein structures–a feature that might induce amyloid formation by edge-to-edge aggregation of beta-sheets [35]. Another example is the 'end-capping' of sequence regions in globular proteins that would otherwise exhibit a high amyloid-forming propensity by charged or structure-disrupting residues [36].

The stabilising effect of disordered flanks is closely related to steric stabilisation of colloids by polymers. Indeed, steric stabilization has been exploited extensively in material and drug design to stop colloids aggregating [37] or to increase the lifetime of hydrophobic drugs by attaching the drug to block copolymers with a hydrophobic middle and hydrophilic flanks [38]. The latter experiments show that steric stabilisation of hydrophobic moieties is highly relevant in biological systems but, as is often the case, evolution "discovered" this effect first.

The present work provides a testable hypothesis for the abundance of disordered regions in proteins: it suggests that disordered flanks adjacent to hydrophobic motifs can suppress aggregation of the hydrophobic peptides in solution. The hypothesis that we put forward gives a basis for in vitro or in vivo experiments into the effect of hydrophilic disordered flanks on the aggregation, solvability and toxicity of hydrophobic peptides. Confirmation of our predictions in a biological context may lead to new methods that could increase the bioavailability of hydrophobic peptides.

Methods

3D Lattice Model

We use a coarse grained representation of a peptide chain where each residue occupies a single point on a cubic lattice [26]. Neighboring residues that would be covalently bound in a peptide chain are required to be on neighbouring lattice sites (Figure 1). Residues interact when residing on neighbouring sites. The internal energy of a configuration is given by:

$$E = \frac{1}{2} \sum_{i,j}^{N} M_{A(i),A(j)} \cdot C_{i,j} \qquad (1)$$

where $A(i)$ gives the amino acid at residue i, $C_{i,j} = 1$ when residues i and j interact and $C_{i,j} = 0$ otherwise. The interaction matrix M gives the pairwise interactions between all 20 amino acids and is based on the occurrence of amino acids in close proximity in experimentally determined protein structures [24,25]. The interaction matrix is normalised with respect to Threonine [25], so that all pairwise interaction energies of Threonine are set to zero. We use this in our simulations to observe the purely entropic contributions of the disordered flanks.

The interaction matrix used here is based on structural proteins, while pairwise interactions in unstructured regions may have slightly different propensities. One may expect that hydrophobic residues in unstructured peptide sequence may be some what less hydrophobic due to the exposed backbone. In this case it may be that the number of hydrophobic residues needed for peptide aggregation is slightly higher than in the current work, but we expect that the qualitative effects of the aggregation remain similar.

Monte Carlo Simulation

We use a Monte Carlo simulation technique where trial steps are accepted according to:

$$P_{acc} = \min\left\{1, \exp\left(\frac{-\Delta E}{k_B T}\right)\right\} \qquad (2)$$

where T is the simulation temperature, k_b is the Boltzmann constant and $-\Delta E$ is the difference in energy between the new and old configuration of the model. Trial moves are either internal moves, changing the configuration of a chain (end move, corner flip, crank shaft, point rotation), or rigid body moves, changing the position of the chain relative to other objects (rotation, translation), see ref. [27] for more details. At each iteration a single local trial move is performed and a global trial move move (including point rotations) is performed with the probability ($P_{global} = 0.1$). In the binding simulations, only rigid body moves are applied to 'rigid' binding motifs, whereas the configurations of the flexible binding motifs are sampled with both internal and rigid body moves.

The volume of the simulation box ($60 \times 60 \times 60$ lattice points) was kept constant, yielding a concentration for the peptide that is higher than that typical of signalling peptides in a cell (approximately 10–1000 times higher). However, the cytosol will contain other signalling peptides that, if not properly protected, could participate in aggregation. Moreover, as argued in the Supplementary Material (Text S1), the peptide solutions in our model are still sufficiently dilute to make it possible to extrapolate our findings to the typical concentrations that prevail inside a cell.

Parallel tempering, or temperature replica exchange, was used to converge more rapidly to sampling of equilibrium configurations. Multiple simulations at different temperatures were run in parallel, while trying to swap temperatures every 50000 moves with 10000 trial temperatures swaps in each simulation. A trial swap between the temperatures of two replicas was accepted with a probability [39–41]:

$$P_{acc} = \min\left\{1, \exp\left(\frac{\Delta E \cdot \Delta 1 / T}{k_B}\right)\right\} \qquad (3)$$

Design of Binding Site

The design of binding interface (i.e. the contacts between the binding motif and the binding groove) was achieved through a Monte Carlo algorithm that interchanges amino acids, while optimising the total energy of the bound state and keeping the variance of the amino acids high, see [27,28] for more details.

Sampling of Configurations

In order to estimate the probability distribution $P(x)$ (where x is an "order parameter", such as C_{ext}, the number of external contacts), we use both configurations of accepted and rejected trial moves weighted by the Boltzman factors of each configuration [42].

The amount of binding of the binding motif to the substrate is tracked by comparing the number of (non-covalent) contacts $C_{i,j}$ in a configuration to the contacts present in the fully bound state $C_{i,j}^{nat} \in \{0,1\}$. Then the total number of native binding contacts is defined as:

$$C_{nat} = \frac{1}{2} \sum_{i,j}^{N} C_{i,j}^{nat} C_{i,j} \qquad (4)$$

where $\mathcal{N}$ is the total number of residues in the binding motif (excluding the flanking regions).

Tracking aggregation of multiple binding motifs is done by considering the total number of external contacts C_{ext}:

$$C_{ext} = \frac{1}{2} \sum_{k,l \neq k}^{M} \sum_{i,j}^{N} C_{k_i,l_j} \qquad (5)$$

where M is the total number of chains in the simulation box and C_{k_i,l_j} is a contact between residue i in chain k and residue j in chain l. Note that Threonine-Threonine contacts do not contribute to C_{ext}.

The amount of binding is given by:

$$P_b = \begin{cases} 1 & \text{if } C_{nat} \geq 0.8 \, max\{C_{nat}\} \\ 0 & \text{otherwise} \end{cases} \qquad (6)$$

The constant volume heat capacity is calculated as:

$$C_v = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2} \qquad (7)$$

Ensemble averages for an order parameter x are given by:

$$\langle x \rangle = \sum_x x P(x) \qquad (8)$$

where $P(x)$ is estimated as before.

Grand Canonical Simulation

A grand canonical Monte Carlo simulation was performed to investigate the aggregation behaviour of binding motifs at a

constant (low) concentration of these peptides. Trial insertions and deletions were performed with a probability of $P_{insert} = P_{delete} = 0.005$ per move. Trial insertion of new chains (with an identical sequence) were accepted with:

$$P_{acc} = \min\left\{1, \frac{V}{N+1}\exp(\mu\beta)\right\} \qquad (9)$$

and deleted with:

$$P_{acc} = \min\left\{1, \frac{N}{V}\exp(-\mu\beta)\right\} \qquad (10)$$

where $\beta = \dfrac{1}{k_B T}$, N is the number of free chains in the simulation box before the move, V is the volume of the box, and μ the chemical potential. The volume was kept constant at $30 \times 30 \times 30$ lattice points and $exp(\mu\beta)$ was kept constant at $3 \cdot 10^{-6}$ in all simulations. A single peptide chain was simulated in a separate box, at the same temperature, to generate new configurations for insertion into the main simulation box. Only free chains were inserted and removed, i.e. no chains that make an external contact with another chain.

Since the chains were simulated at very low density, moves are likely that remove the only peptide chain from the simulation box. At such an event the number of trial insertion moves (M_i) to re-entrance was taken as:

$$M_i = \frac{\ln(U)}{\ln(1 - V\exp(\mu\beta))} \qquad (11)$$

where U is a random, uniformly distributed variable on the interval [0,1].

The total number of sampling steps is given by the total number of trial moves (S):

$$M = \left(1.0 + P_{global}\right)\frac{M_i}{P_{insert}} \qquad (12)$$

The order parameters and internal energy are all zero for the empty simulation box.

Images

Images in Figures 1 and 4 were produced using the UCSF Chimera package [43].

Supporting Information

Figure S1 Binding motifs embedded in different environments bound to the same substrate From left to right: (A) a binding motif, (B) a binding embedded in disordered flanks and (C) a binding motif in a rigid structure. The yellow residues are hydrophobic, the blue negatively charged, the red positively charged and the grey hydrophilic.
Found at: doi:10.1371/journal.pcbi.1000241.s001 (0.15 MB PNG)

Figure S2 Unbound binding motifs From left to right: (A) a binding motif, (B) a binding embedded in disordered flanks and (C) a binding motif in a rigid structure. The yellow residues are hydrophobic, the blue negatively charged, the red positively charged and the grey hydrophilic.
Found at: doi:10.1371/journal.pcbi.1000241.s002 (0.06 MB PNG)

Text S1
Found at: doi:10.1371/journal.pcbi.1000241.s003 (0.09 MB PDF)

Author Contributions

Conceived and designed the experiments: SA DF. Performed the experiments: SA. Analyzed the data: SA DF. Wrote the paper: SA DF.

References

1. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293: 321–331.
2. Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, et al. (1998) Thousands of proteins likely to have long disordered regions. Pac Symp Biocomput. pp 437–448.
3. Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, et al. (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput. pp 473–484.
4. Garner E, Cannon P, Romero P, Obradovic Z, Dunker AK (1998) Predicting disordered regions from amino acid sequence: common themes despite differing structural characterization. Genome Inform Ser Workshop Genome Inform 9: 201–213.
5. Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, et al. (2005) Comparing and combining predictors of mostly disordered proteins. Biochemistry 44: 1989–2000.
6. Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41: 415–427.
7. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ (2000) Intrinsic protein disorder in complete genomes. Genome Inform Ser Workshop Genome Inform 11: 161–171.
8. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 337: 635–645.
9. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, et al. (2001) Intrinsically disordered protein. J Mol Graph Model 19: 26–59.
10. Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27: 527–533.
11. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 323: 573–584.
12. Fink AL (2005) Natively unfolded proteins. Curr Opin Struct Biol 15: 35–41.
13. Uversky VN, Oldfield CJ, Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit 18: 343–384.
14. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. the roles of intrinsic disorder in protein interaction networks. FEBS J 272: 5129–5148.
15. Gunasekaran K, Tsai CJ, Nussinov R (2004) Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. J Mol Biol 341: 1327–1341.
16. Coluzza I, Frenkel D (2007) Monte carlo study of substrate-induced folding and refolding of lattice proteins. Biophys J 92: 1150–1156.
17. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, et al. (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics 9(Suppl 1): S1.
18. Mészáros B, Tompa P, Simon I, Dosztányi Z (2007) Molecular principles of the interactions of disordered proteins. J Mol Biol 372: 549–561.
19. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, et al. (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 44: 12454–12470.
20. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, et al. (2006) Analysis of molecular recognition features (morfs). J Mol Biol 362: 1043–1059.
21. Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, et al. (2007) Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. Biochemistry 46: 13468–13477.
22. Vacic V, Oldfield CJ, Mohan A, Radivojac P, Cortese MS, et al. (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. J Proteome Res 6: 2351–2366.
23. Fuxreiter M, Tompa P, Simon I (2007) Local structural disorder imparts plasticity on linear motifs. Bioinformatics 23: 950–956.

24. Miyazawa S, Jernigan RL (1993) A new substitution matrix for protein sequence searches based on contact frequencies in protein structures. Protein Eng 6: 267–278.

25. Betancourt MR, Thirumalai D (1999) Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. Protein Sci 8: 361–369.

26. Shakhnovich EI, Gutin AM (1993) Engineering of stable and fast-folding sequences of model proteins. Proc Natl Acad Sci U S A 90: 7195–7199.

27. Coluzza I, Muller HG, Frenkel D (2003) Designing refoldable model molecules. Phys Rev E Stat Nonlin Soft Matter Phys 68: 046703.

28. Coluzza I, Frenkel D (2004) Designing specificity of protein-substrate interactions. Phys Rev E Stat Nonlin Soft Matter Phys 70: 051917.

29. Privalov PL, Tiktopulo EI, Venyaminov SY, Griko YV, Makhatadze GI, et al. (1989) Heat capacity and conformation of proteins in the denatured state. J Mol Biol 205: 737–750.

30. Borrero EE, Escobedo FA (2006) Folding kinetics of a lattice protein via a forward flux sampling approach. J Chem Phys 125: 164904.

31. Oma Y, Kino Y, Sasagawa N, Ishiura S (2005) Comparative analysis of the cytotoxicity of homopolymeric amino acids. Biochim Biophys Acta 1748: 174–179.

32. Baglioni S, Casamenti F, Bucciantini M, Luheshi LM, Taddei N, et al. (2006) Prefibrillar amyloid aggregates could be generic toxins in higher organisms. J Neurosci 26: 8160–8167.

33. Cheon M, Chang I, Mohanty S, Luheshi LM, Dobson CM, et al. (2007) Structural reorganisation and potential toxicity of oligomeric species formed during the assembly of amyloid fibrils. PLoS Comput Biol 3: e173. doi:10.1371/journal.pcbi.0030173.

34. Monsellier E, Chiti F (2007) Prevention of amyloidlike aggregation as a driving force of protein evolution. EMBO Rep 8: 737–742.

35. Richardson JS, Richardson DC (2002) Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc Natl Acad Sci U S A 99: 2754–2759.

36. Rousseau F, Serrano L, Schymkowitz JW (2006) How evolutionary pressure against protein aggregation shaped chaperone specificity. J Mol Biol 355: 1037–1047.

37. Elbert DL, Hubbell JA (1998) Self-assembly and steric stabilization at heterogeneous, biological surfaces using adsorbing block copolymers. Chem Biol 5: 177–183.

38. Kataoka K, Harada A, Nagasaki Y (2001) Block copolymer micelles for drug delivery: design, characterization and biological significance. Adv Drug Deliv Rev 47: 113–131.

39. Lyubartsev AP, Martsinovski AA, Shevkunov SV, Vorontsov-Velyaminov PN (1992) New approach to Monte Carlo calculation of the free energy: method of expanded ensembles. J Chem Phys 96: 1776–1783.

40. Marinari E, Parisi G (1992) Simulated tempering: A new monte carlo scheme. Europhys Lett 19: 451–458.

41. Geyer CJ, Thompson EA (1995) Annealing Markov Chain Monte Carlo with applications to ancestral inference. J Am Stat Assoc 90: 909–920.

42. Frenkel D (2004) Speed-up of Monte Carlo simulations by sampling of rejected states. Proc Natl Acad Sci U S A 101: 17571–17575.

43. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF chimera—a visualization system for exploratory research and analysis. J Comput Chem 25: 1605–1612.