doi:10.1016/j.jmb.2006.12.006 J. Mol. Biol. (2007) 367, 262–274

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Induced Fit, Folding, and Recognition of the NF-κB-Nuclear Localization Signals by IκBα and IκBβ

Joachim Lätzer[1,2], Garegin A. Papoian[3], Michael C. Prentiss[1,2] Elizabeth A. Komives[1] and Peter G. Wolynes[1,2,4]*

[1]*Department of Chemistry and Biochemistry, University of California at San Diego 9500 Gilman Drive, La Jolla CA 92093-0371, USA*

[2]*Center for Theoretical Biological Physics, University of California at San Diego 9500 Gilman Drive, La Jolla CA 92093, USA*

[3]*Department of Chemistry University of North Carolina Chapel Hill, North Carolina NC 27599, USA*

[4]*Department of Physics University of California at San Diego, 9500 Gilman Drive La Jolla, CA 92093, USA*

**Corresponding author*

Protein structure prediction codes based on the associative memory Hamiltonian were used to probe the binding modes between the nuclear localization signal (NLS) polypeptide of NF-κB and the inhibitors IκBα and IκBβ. Experimentally, it is known that the NLS polypeptide is unstructured in the NF-κB complex with DNA but it forms an extended helical structure with the NLS (residues 301–304) between the two helices in the NF-κB/IκBα complex. The simulations included the NF-κB(p65) and (p50) NLS polypeptides and various mutants alone and in the presence of IκBα and IκBβ. The simulations predict that the NLS polypeptide by itself binds tightly to IκBα and IκBβ. In the NF-κB (p50/p65) heterodimer, the p50 NLS is predicted to remain free to bind to importin α. In the interaction with IκBα, both p65 NLSs are predicted to be bound. In IκBβ, the NLS polypeptide binds to two binding sites, as seen in the crystal structure, with one site heavily favored for stable binding.

Keywords: protein folding; protein binding; induced fit; energy landscape; protein recognition

Introduction

The import selectivity of nuclear proteins from the cytosol relies on nuclear localization signals (NLS). These generally are short sequences of 3–20 amino acid residues, normally rich in lysine and arginine, which bind to nuclear import receptors.[1,2] Monopartite NLSs contain only a single cluster of positively charged residues, and bipartite NLSs contain two stretches of basic residues connected typically by a 10–12 residue linker. The prototypical monopartite NLS is the simian virus 40 (SV40) large T antigen (TAg) NLS of sequence PKKKRKV.[3] This NLS sequence is very specific and mutation of a single residue, K128, leads to loss of binding to the nuclear import factor resulting in cytoplasmic retention of TAg.[4] A consensus sequence that represents the diversity of NLSs is K-(K/R)-X-(K/R), where X is any amino acid.[5] This consensus sequence must be recognized by the nuclear import receptor.

For the NLS to be active, it needs to be exposed for binding to the surface of the nuclear protein–import receptor complex. IκB, the inhibitor of NF-κB, deactivates the NF-κB nuclear localization signal by physically masking it.[6,7] The interactions of the different NLSs of the NF-κB family members, homo- or heterodimers made from five subunits, have evolved to achieve very specific recognition and binding to members of the IκB-family. For example, IκBα, the most abundant IκB, binds and inhibits NF-κB p65 homodimers but not p50 homodimers.[8] Part of the reason for this interaction specificity lies in the recognition of the NF-κB NLS polypeptide. The NLS polypeptides of p50 and p65 are defined as the NLS-containing carboxy-terminal fragments with lowest sequence homology within the rel homology do-

Abbreviations used: NLS, nuclear localization signal; Tag, large T antigen; AMH, associative memory Hamiltonian; CE, combinatorial extension; RMSD, root-mean-square deviation.

E-mail address of the corresponding author: pwolynes@chem.ucsd.edu

0022-2836/$ - see front matter © 2007 Published by Elsevier Ltd.

main of the NF-κB family.[9] The specificity of IκBα for p65 comes partly from the fact that the p50 NLS does not bind to IκBα with significant affinity.[10] Comparison of the crystal structures of p50/p65 complexed to IκBα and to DNA show electron density for the p65 NLS polypeptide when bound to IκBα but not when bound to DNA.[11,12] This result suggests that the p65 NLS polypeptide is flexible in the unbound state and becomes more ordered upon forming a complex with IκBα.[10]

The advantage of the NLS being inherently flexible is that local structure can be modified in response to different molecular targets. This allows competitive binding to several different targets, affording the necessary non-linearity of a control circuit.[13] NLSs bound to importin α usually adopt extended structures.[14] The p65 NLS polypeptide, residues 289–320 of the rel homology domain of p65, binds to IκBα in a split helical conformation.[12,15] Thus, the disordered structure of the NLS polypeptide when NF-κB is free or bound to DNA allows it to recognize either importin α or the various IκB isoforms.

In order to probe how the NF-κB NLS polypeptides achieve flexibility and specificity at the molecular level, we performed simulations to predict structures using the optimized associative memory Hamiltonian (AMH) method for the free and IκB-bound NLS polypeptides.[16] We note that the word predict can lead to misunderstanding, since the crystal structures of IκBα with NF-κB(p50/p65) and of IκBβ with NF-κB(p65/p65) are already known and deposited with the RCSB Protein Data Bank.[15,17] Nevertheless, we show here that our energy function, without explicit knowledge of the native structure, can capture these dominant binding modes for the full length proteins correctly, and thus predict them, and that the computational analysis allows us to elucidate how evolution has led to the necessary binding specificity. We analyze the effects by simulating shorter constructs that give insight into the individual roles of the binding partners in the binding process. The results from these simulations show that the free NLS polypeptide is thermodynamically guided to adopt a helix-turn-helix structure with the NLS itself forming the turn. Simulations of the IκB-bound NLS polypeptides show that the p50 NLS polypeptide does not interact specifically with IκBα, while the p65 NLS polypeptide is predicted to have several distinct binding modes, with the NF-κB(p50/p65)-IκBα crystal-structure-like conformation being one of them.

Simulation of the p65 NLS polypeptide interacting with IκBβ reveals two conformations, as found in the NF-κB(p65/p65)-IκBβ crystal structure. The simulations therefore make a new prediction; when the structure of NF-κB(p65/p65) bound to IκBα is determined, both NLS polypeptides will be bound on opposite faces of the IκBα.

The simulations with IκBα uncover a beautifully clear example of “induced fit”, arguing for greater specificity and provide a rationale for nature’s design scheme for NF-κB NLS polypeptides.[18,19] The specific basic residues of the NF-κB NLS both interrupt the helical propensity of the signal and form crucial contacts with IκBα/β, bringing the helical portions into position to "cap" the ankyrin repeat domain.

Results

Validation and benchmarking of the AMH method

This study of the IκB-NF-κB system aims to elucidate the folding and binding of a 30 residue long helical polypeptide, the NLS polypeptide, which becomes structured in the vicinity of its binding partner IκBα. In general, predicting folding and binding structures is a challenging problem when no experimental data are available. Here, we used molecular dynamics simulations with AMH as the energy function for identifying the binding interface, while folding the NLS.[16] This search problem is much simpler than total *ab initio* prediction, since the conformation of the binding partner, IκBα, is kept native-like but is allowed to fluctuate around the native basin with thermal energy k_BT. The current study is then reduced to predicting the folding of the NLS onto a fixed protein surface. This constrained search problem is much easier than predicting folding and binding of two large proteins of unknown structure entirely from scratch. The inter and intra-residue interactions between the NLS and the binding partner IκBα are physically equivalent. Many meaningful protein dimer studies, such as Gô-model studies, assign the same contact energies to inter and intra-residue contacts.[20] The AMH approximates a physically and chemically correct energy function and should, in principle, be appropriate to describe the complete folding and binding process of this simplified problem.[16] Water effects, which are important in protein binding, are only implicitly incorporated in the AMH. The NLS/IκBα binding interface, which has mostly hydrophobic native contacts between the NLS and IκBα was found to be adequately described by AMH.

Several different binding and folding scenarios were simulated in the present work. In the first scenario, we present the folding of the NLS polypeptide onto the IκBα ankyrin repeat. Hydrogen/deuterium exchange experiments performed in the Komives laboratory show that the first three ankyrin repeats are protected and hence folded stably.[21] It has been established experimentally that the NLS polypeptide binds to these ankyrin repeats.[22] Here, we show that the present computation can reproduce the crystal structure of the NLS bound to IκBα when the two binding partners are connected by a glycine linker. This is not an unreasonable construct to study, as biology itself has these two parts connected by a glycine linker in

the NF-κB precursor protein. We have varied the length of the glycine linker to confirm that the results are independent of linker length.

To understand the reliability of this approach, we studied whether we could successfully identify binding modes of comparable fragments in other systems. First, we constructed a shortened endonuclease dimer (PDB code 1m0i), where the two dimers were connected by a glycine linker. Eighty AMH annealing runs of the endonuclease with 75% of the protein constrained to be native-like resulted in only 17 structures misfolding, and all of the remaining structures representing the correct binding site after clustering using the Fitch-Margoliash method to generate a phylogenetic tree from the distance map, where distance is a measure of similarity (Figure 1(a)). The binding interface, as well as the fold of the simulated structures, resembles the crystal structure conformation found in the PDB. The AMH overfolds the endonuclear helix (green) and aligns it with a slight tilt when compared to the crystal structure (blue). The interface between the two endonucleases is predicted correctly with high overlap of native inter-residue contacts. The combinatorial extension (CE) algorithm calculates pairwise structure alignments. The CE *Z*-score provides a measure of statistical significance of the alignment, and structures with a *Z*-score of 3.5 or higher are generally judged to have a similar fold comparable to that obtained in a typical homology model.[23] The *Z*-score between individual structures obtained from the molecular dynamics simulations are in the range of 3.3–.7, again showing a great deal of similarity in overall topology and conservation of the helical secondary structural elements. The average root-mean-square deviations (RMSDs) from the crystal structure of the dimer were 2.3 Å for the fragment only and 3.2 Å for the fragment and the interface contact residues, respectively.

Similarly, structures predicted for the folding of the N-terminal helix of 30 residues of myoglobin (PDB code 104m) when the rest of the molecule is constrained also showed profound topological agreement (Figure 1(b)). In this case, first the helix was completely unfolded and allowed to dock anywhere on the protein surface of myoglobin and then 80 annealing runs were performed and the final structures were clustered. All structures obtained in these simulations docked at the correct binding site. The CE *Z*-score for these structures were in the range of 3.8–4.3. The RMSD from the crystal structure was 0.9 Å for the fragment and 1.5 Å for the fragment and the interface residues. These two test cases validate the present AMH approach for predicting binding conformations of modestly sized fragments like the NLS to pre-formed protein structures.

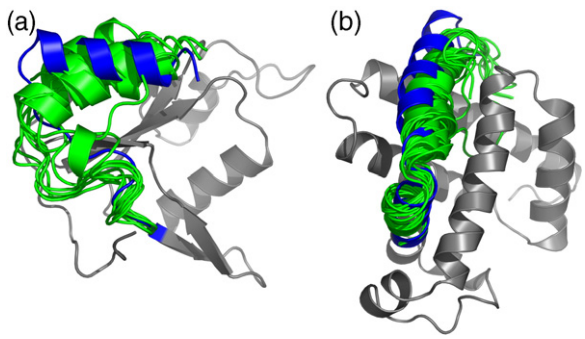

Figure 1. (a) Overlaid structures of endonuclease from simulation and experiment. Here, the restrained part is shown in grey, the crystal structure is shown in blue and the structures obtained with the AMH are shown in green. (b) Overlay of myoglobin structures from simulation and experiment. The coloring is the same as in (a).

Predicted structure of the p50, p65, and nucleoplasmin NLS polypeptides

In order to obtain an overview of the possible structures that the free NLSs could assume when their binding partners are not present, we predicted the conformations of the free polypeptides using AMH. This study was motivated by the fact that crystallography shows the NLS polypeptide is disordered when p65 is bound to DNA with no interpretable electron density of the NLS polypeptide residues.[11] Annealing runs carried out on the p65 NLS polypeptide (PDTDDRHRIEEKRKRTYETFKSIMK) alone indicated one family of structures with two helices that formed close contacts, connected by a loop (Figure 2). The first proline initiates the first helix of about five residues; Asp291–Arg295. The NLS (Lys_{301}ArgLysArg$_{304}$) was found in a bend and turn region. A second helix usually spanned from Glu307 to Met313. The helices formed close contacts with a separation of about 7 Å, with a predicted end-to-end-distance of about 5 Å. The *Q*-score of 0.43, which measures the similarity of the annealed structures relative to each other, also confirmed that the structures were similar to each other (Figure 2(a)). The annealed structures showed somewhat different orientations of the non-helical termini but, excluding the last two residues each of the N and C termini did result in *Q*-scores higher than 0.5. The *Z*-score between individual structures obtained from the molecular dynamics simulations are in the range of 3.3–3.7, again showing a great deal of similarity in overall topology. The secondary structure of the helical fragments was conserved. The average RMSDs for the first helix (residues 294–302) and the second helix (residues 305–314) from the NLS polypeptide in the crystal structure, were 0.87 Å and 0.64 Å, respectively.

Structures for the p50 NLS (PLYYPEIKDKEEVQRKRQKLMPNFSDSFGGGSGAG) and the nucleoplasmin NLS (AVKRPAATKKAGQAKKKKL) that were obtained in annealing runs following the same protocol as for the p65 NLS are also shown in Figure 2. All annealed structures of the p50 NLS were

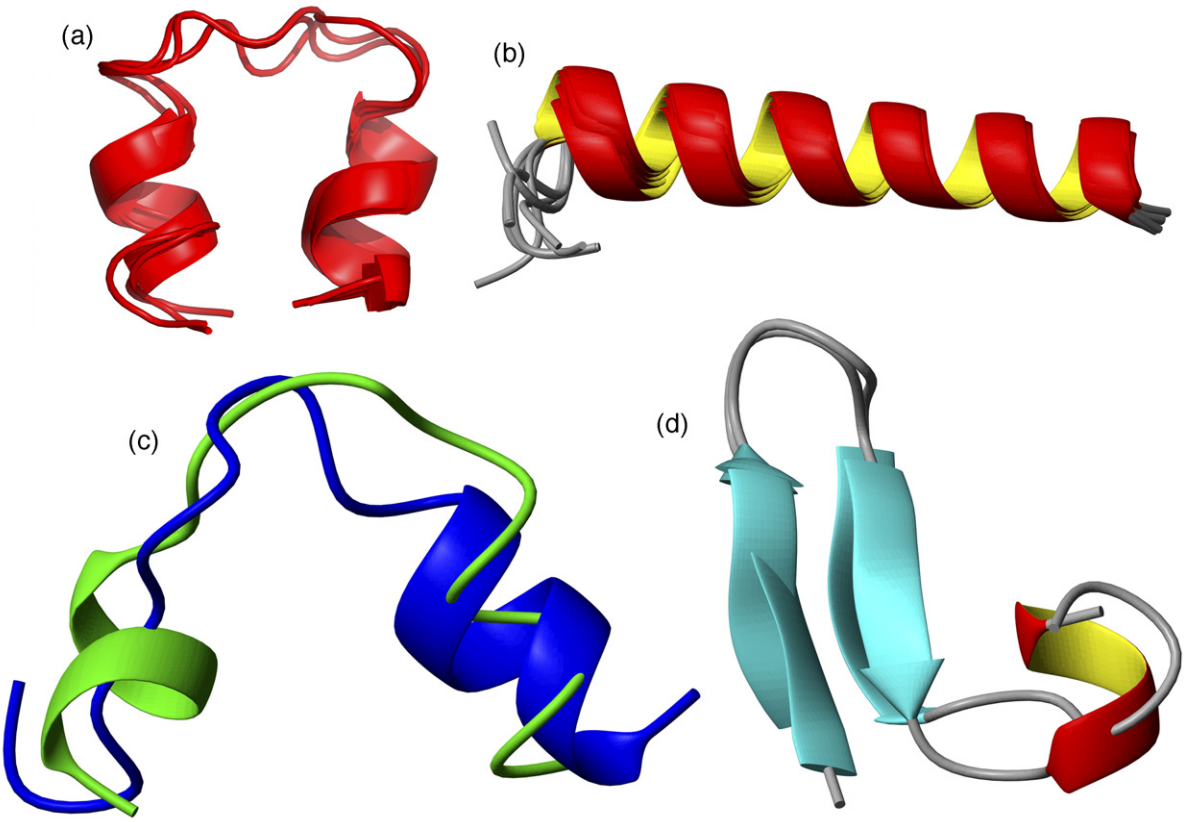

Figure 2. (a) Overlaid structures obtained in annealing runs to predict the structure of the p65 NLS polypeptide by itself. Two helices are connected by a breaking or kink region that contains the basic NLS residues. The structures are compact rather than elongated. (b) Structures obtained for the p50 NLS sequence folded uniquely into one straight α-helix. (c) and (d) The nucleoplasmin sequence showed frustrated folding, resulting in many structures with different folds and secondary structure content.

structurally equivalent to each other. Structural order parameters such as the CE *Z*-score indicate that all the predicted structures of the p50 NLS share the same fold, a long α-helix. This is interesting, because the sequence of the p50 NLS, although different from the p65 NLS, also has a basic central region, but in this case a break in the middle was not observed. The annealing runs suggested that the energy landscape of folding for the nucleoplasmin NLS sequence is more frustrated. The resulting structures vary tremendously in topology, some showing all-α secondary structures as well as α-β secondary structures. The simulations showing that the nucleoplasmin NLS adopts an extended structure agree with experiments, where the nucleoplasmin NLS binds to the import factor importin in an extended structure with hydrogen bonding interactions.[14]

Folding of the IκBα-NLS construct

Multiple (130) annealing runs were performed using the AMH algorithm to obtain structures of the IκBα-NLS construct linked by nine glycine residues. The linker assures proximity of the NLS to IκBα in the simulation. The use of a glycine linker is biologically justified for the NF-κB/IκB system, because the precursor p105 in fact contains NF-κB and IκB linked by a glycine-rich linker region.[24] Structures with various degrees of similarity to the crystal structure were obtained from these simulations. The Fitch-Margoliash method was used to cluster these structures according to their similarity as measured in distance, *d*, and a phylogenetic tree was created from the distance map. This phylogenetic tree showed four main clusters of structures with low relative *d*, and hence significant similarity. The high relative *Q*-scores within each cluster also indicate similar global folds that represent distinct structural families or specific binding modes. The center of each branch was chosen as the representative structure for that individual cluster (Figure 3). Cluster 4, the dominant cluster with a third of the structures, is composed of those conformations that most resembled the X-ray crystal structure based on the *Q*-score that included contacts of the NLS polypeptide along with contributions from the interface of the NLS polypeptide with IκBα. Clusters 1, 2 and 3 all had lower *Q*-scores of about 0.25 relative to the crystal structure.

Cluster 4 had a *Q*-score of 0.42 of the basin center with respect to the crystal structure (Figure 3(d)). The cluster had an overall RMSD of 0.33(±0.05) Å showing very high overlap between structures. The

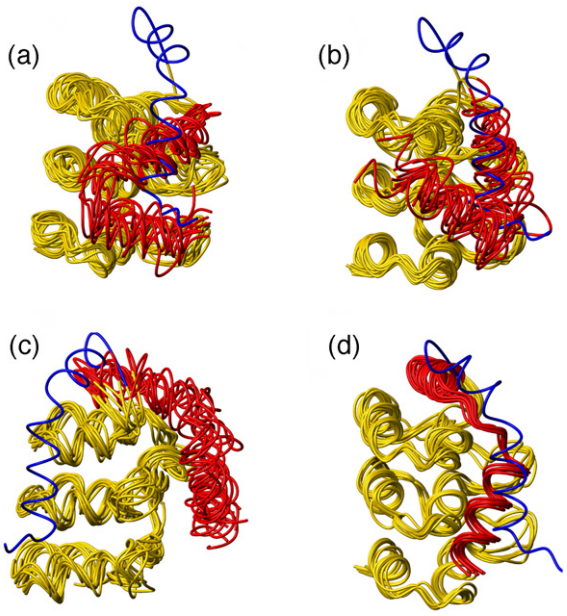

Figure 3. A phylogenetic tree was created for the structures obtained in the annealing runs of the IκBα-p65 NLS construct using the Fitch-Margoliash algorithm. The tree showed clustering in (a)–(d) four main groups, which were analyzed as possible binding modes. The inhibitor IκBα is colored yellow and the p65 NLS polypeptide is colored red. For comparison, the crystal structure conformation is shown in blue. (a) and (b) Clusters 1 and 2 correspond to overly compact structures that folded and then docked onto the protein surface of IκBα. (c) The cluster 3 (C) structures form a basin of "symmetric" structures, in which the NLS polypeptide docked at an alternative, biologically possible, protein surface of the inhibitor. (d) The cluster 4 structures are structurally close to those found in the X-ray crystal structure.[12]

first helix (residues 294–302) and the second helix (residues 305–314), of the clustered structures had even lower RMSDs, 0.19(±0.14) Å and 0.13(±0.03) Å, respectively. Further, the CE *Z*-scores of the individual members of the cluster relative to the cluster center were about 5, indicating essentially identical folds and justifying the use of the cluster center as representing the structure of the cluster. Comparison of the structure from the center of basin 4 relative to the X-ray crystal structure gave an overall RMSD of the NLS residues along with the interface residues (see Materials and Methods) of 3.1 Å (see also Table 1). The secondary structure elements were native-like, with an RMSD of 0.57 Å for the first helix and 0.34 Å for the second helix. Similar native-like features of the structures were obvious. These included the observation that the second helix caps the top of IκBα and the first helix binds appropriately to the hydrophobic fingers of IκBα. The RMSDs from the crystal structure for the structures in basin 4 are smaller than the fluctuations inferred from the measured *B*-factors (PDB file 1NFI) for residues 293–302. Thus, the structures predicted in this basin are consistent with observation. In the simulations, the second helix packed somewhat more closely to the hydrophobic top of IκBα than would appear to be the case in the crystal structure (Figure 3(d)).

Clusters 1–3 had the same *Q*-score to the native structure, but were distinct clusters having less similarity to each other. When only atoms from the NLS were used to calculate the RMSD from the crystal structure, the results for clusters 1–3 were 4.9 Å, 5.9 Å and 3.4 Å, respectively. The RMSDs of the two helical fragments of basin 3 structures from the corresponding crystal structure fragments were 1.1 Å and 2.2 Å, respectively. Thus, the basin 3 NLS had a fold similar to that of the crystal structure; however, when the interface residues were included in the calculation, the RMSD was 7.5 Å, indicating clearly that basin 3 structures (Figure 3(c)) bind differently from the mode observed in the crystal structure. The NLS still binds to the IκBα β-hairpin fingers, and also still caps the hydrophobic top; however, the NLS polypeptide binds on the other side of the β-hairpins of the ankyrin repeat domain than was seen in the crystal structure.

Clusters 1 and 2 constituted about 55% of the observed structures. Neither of these clusters was as homogeneous as cluster 4 in terms of CE *Z*-score; however, the secondary structure of the helical fragments was preserved. The structures belonging to these clusters exhibited no capping of IκBα by the NLS polypeptide (Figure 3(a) and (b)). In the annealing simulations, structures arriving at these binding modes appeared to fold first to an overly compact structure resembling the folding found for the unbound NLS polypeptide, and this configuration later docked onto the protein surface of IκBα.

Important contacts between the NLS polypeptide and IκBα

Contact maps of the four clusters were constructed to identify the most important residues involved in the NLS recognition of IκBα (Table 2). As expected, the cluster 4 structures formed the largest number of native contacts between IκBα and the NLS polypeptide, while clusters 1 and 2 had only a few residues with high native contact probability, and the structures in cluster 3 showed no native contact formed at all. Clusters 1 and 2 contained mainly non-native contacts formed between helix 2 (residues 305–314), and IκBα. In these binding

Table 1. Summary of RMSD for simulated constructs

			RMSD	
Proteins	Structure	NLS	Interface	NLS + interface
p65NLS+ IκBα	Basin 1	4.9	5.9	6.4
p65NLS+ IκBα	Basin 2	5.9	6.6	7.2
p65NLS+ IκBα	Basin 3	3.4	6.7	6.2
p65NLS+ IκBα	Basin 4	1.7	2.1	2.1
p65NLS+ IκBα+p65 remainder	Basin 1	2.2	2.4	2.8
p65NLS+ IκBα+p65 remainder	Basin 2	3.5	18.3	22.4
p65NLS+ IκBβ	Basin 1	1.7	1.2	2.1
p65NLS+ IκBβ	Basin 2	4.6	5.2	5.4

Table 2. Important contacts for basins 1–4

	Native contacts		Non-native contacts	
	NLS	IκBα	NLS	IκBα
Basin 1	Ile298	Ile120	Phe309	Ile83
			Phe309	Ile117
			Phe309	Leu120
			Met313	His79
			Met313	Leu80
			Met313	Ile83
Basin 2	Ile298	Ile120	Phe309	Ile83
			Met313	Leu80
			Met313	Ile83
Basin 3			Tyr306	Phe87
			Phe309	Phe77
Basin 4	ASP294	Ile120	Glu 299	His 84
	Arg295	Ile120	Lys301	Ile112
	Arg295	Thr121	Arg302	Leu80
	Ile298	Ile83	Tyr306	Leu80
	Ile298	Leu117	Phe309	Phe77
	Ile298	Ile120	Phe309	Phe103
	Arg302	Ile83	Ile312	Phe103
	Phe309	Leu80	Met313	Leu78
			Met313	Val93

modes, many contacts were observed between helix 2 and IκBα, whereas helix 1 and the NLS made no contacts with IκBα at all, except for the native contact between Ile298 of p65 and Ile120 of IκBα. Highly probable non-native contacts involved interactions of the very hydrophobic residues Phe309 and Met313 with the also very hydrophobic residues Leu80 and Ile83 found in the first ankyrin repeat, and with Leu117 and Ile120 in the second ankyrin repeat. Cluster 3 contact maps showed that Phe77 in the first ankyrin repeat very often interacts strongly with residues Tyr306 and Phe309, even though these interactions are not native contacts.

Cluster 4 structures showed a high probability of forming native contacts only in helix 1. The strong native contacts of helix 1 residues Ile298 and Arg302 were complemented by non-native contacts of the specific NLS residues, Lys301 to Gln112 and Arg302 to Leu80. Additionally, Glu299 formed a non-native contact with His84 with high probability. An important residue in helix 2 is Phe309, which forms a native contact to Leu80 as well as non-native contacts to Phe77 and Ile94, while Met313 was found to form non-native contacts to Leu78 and Val93. On average, the NLS formed about 43% of its native interfacial contacts with IκBα. The interfacial RMSD from the crystal structure for the residue–residue pairs of the cluster 4 structures that have backbone heavy-atoms within 5 Å of each other, are in the range of 3–4 Å. This would be considered an acceptable result using the CAPRI criterion for identifying binding surfaces.[25]

Folding of the IκBα-NLS-p65 construct

In the constructs studied so far, the NLS polypeptide could associate freely with IκBα. Several possible binding modes for the free NLS polypeptide were found. One of these is the native binding mode found in the crystal structure. On inspection, the alternative binding modes appeared sterically incompatible with the presence of the remainder of the p65 molecule. To confirm this explicitly, we simulated constructs that contain the remaining p65 residues and limit the geometrical space accessible to IκBα. The simulation runs with the IκBα-NLS-p65 construct, in which the relative geometry of p65 and IκBα was constrained to be like that in the crystal structure, exhibited two clusters of structures. The dominant cluster (65% of structures) resembles the binding mode found in the crystal structure (Figure 4) and it appeared that the p65 helped the N-terminal helix of the NLS find its proper location. Besides yielding the X-ray crystal structure-like conformation, the simulations yielded the second cluster (35% of structures) of conformations in which the fragment of the p65 NLS binds to the main body of the p65 molecule (Figure 4). Thus, when the p65 dimerization domain is present, structures with an overly compact NLS polypeptide are no longer observed, although these were found in the simulations of the smaller construct. The structures of the self-interacting cluster had an RMSD from the crystal structure of 3.5 Å for the NLS only, showing that the NLS adopted a similar extended structure. However, when the interface residues were included, the RMSD from the crystal structure was 22.4 Å stemming from the fact that the NLS is bound to p65, not to IκBα. No basin 1, 2, or 3 structures was observed, presumably because the remainder of the p65 molecule geometrically limited the binding modes to either those resembling the crystal structure, or a self-interacting mode.

IκBα interactions with the nucleoplasmin NLS

To investigate whether the binding of the NLS polypeptide to IκBα is specific or whether other NLSs could function similarly, simulations of the

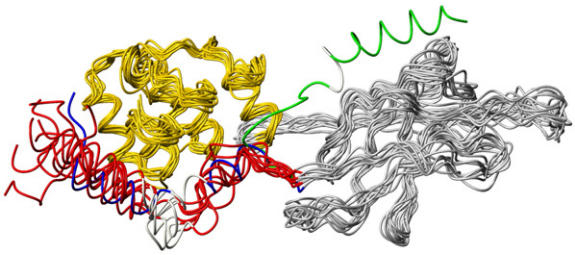

Figure 4. Results from simulation of IκBα-NLS-p65 show geometrical restriction of the possible binding modes. The dimerization domain of the NF-κB p65 (grey) and the IκBα (yellow), were kept fixed in the simulations. The structure of the NLS polypeptide from the crystal structure of the complex of IκBα with NF-κB is colored blue. The main cluster of structures obtained in molecular dynamics simulations with the AMH as energy function reproduced the native structure well (red). The light-grey region in the NLS polypeptide indicates the basic NLS residues, which form a break between the two helices. Some structures were formed in which the NLS polypeptide binds to the p65 dimerization domain. The center of this cluster is colored green.

p65-IκBα construct were performed in which the p65 NLS polypeptide was replaced by the NLS polypeptide of nucleoplasmin, a molecular chaperone whose major function is involved in the assembly of nucleosomes. The structure of the nucleoplasmin NLS polypeptide bound to importin α has been solved.[14] It is of comparable size to the p65 NLS polypeptide but the level of sequence identity is only 6.2%. The AMH simulations of the nucleoplasmin NLS tethered to IκBα and p65 yielded structures in which the NLS polypeptide did not interact with either IκBα or with p65 (data not shown). Thus, AMH simulations suggest that the basic stretch of residues, which has traditionally been the minimal description of the NLS, is insufficient for binding in the IκB/NF-κB system.

Specific effects of the basic NLS residues on IκBα recognition

The basic NLS residues (301–304) were seen to break the helical secondary structure of the p65 NLS in all simulations. Since these residues are required for importin α binding, we sought to define their role in binding to IκBα. Simulated annealing runs were performed on five different IκBα-p65-NLS constructs in which there were alanine substitutions in the NLS. Single mutants; K301A, R302A, K303A and R304A and the quadruple mutant with all four alanine mutations in the NLS were studied. The structures obtained in simulated annealing runs of the K301A and R302A single mutants retained almost wild type amounts (55%) of helical secondary structure. The K303A and R304A mutants showed increased helical contents of up to 77% and formation of one continuous helix with a high binding propensity towards the dimerization domain of p65. The K301A mutant formed mainly long helices in the simulated annealing runs, but the R302A mutant showed a kinked region similar to the crystal structure. None of the mutants produced a structure with native-like topology, although some structures did bind to IκBα. The contact map analysis shows that mutation of R302 to Ala disrupts important native contacts of the first helix to IκBα, perhaps indicating that contacts between the first helix and IκBα are essential for correct complex formation. (Figure 5). To test the idea that the break in the middle of the helix is important, the R304A mutant was made as a peptide spanning residues 289–321 of NF-κB (p65) as well as in the full-length p65. Binding of the peptide was tested by isothermal titration calorimetry and binding of the full-length protein by Biacore. In both experiments, the R304A mutation decreased the binding affinity by twofold and, interestingly, the DCp was much lower for the R304A mutant than for wild type perhaps indicating a change in structure formation during binding (S. Bergqvist, unpublished data).

Binding of the p65 NLS polypeptide to IκBβ

Biochemically, IκBα and IκBβ share the same function and are able to replace each other *in vivo*.[26]

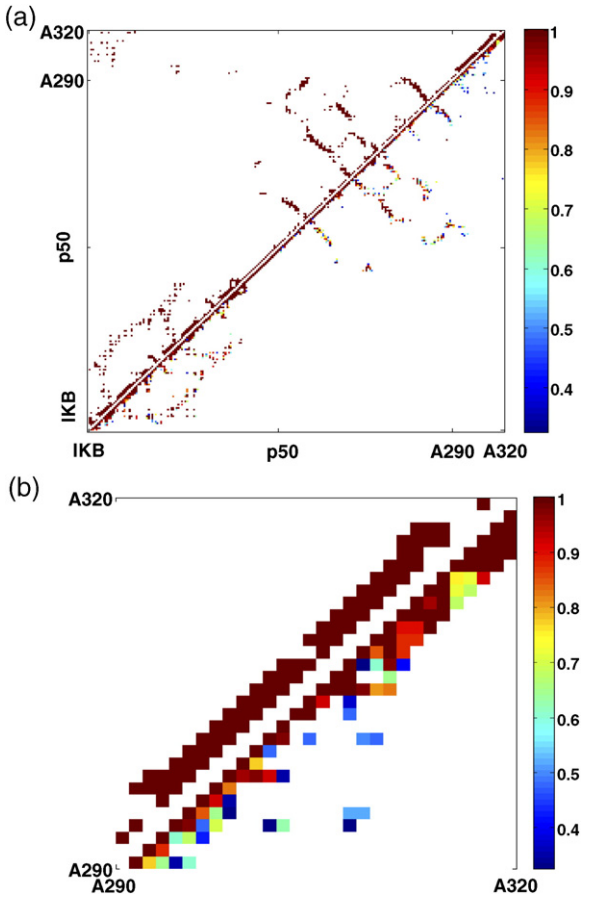

Figure 5. Alanine mutations in the p65 NLS polypeptide were used to probe the role of individual residues in folding and binding of the p65 NLS polypeptide to IκBα. The contact map of the structures obtained in the simulation runs (shown below diagonal) of the Lys302Ala mutant are shown and can be compared to the contact map of the crystal structure (shown above diagonal). Contact probabilities are colored red, indicating a contact that is always formed in the ensemble, while dark blue means that the contact is almost never formed. The complete contact map of the NLS, (a) the p50 remainder and the inhibitor as well as (b) the NLS only are shown.

The crystal structure of IκBβ bound to NF-κB(p65/p65) shows a similar binding site provided by IκBβ as compared to the binding site provided by IκBα in the complex with NF-κB(p50/p65). A second binding site is observed also in the crystal structure of IκBβ bound to NF-κB(p65/p65) with weaker electron density.[27]

AMH simulations of the 30 residue p65 NLS polypeptide interacting with IκBβ yielded only one main basin of structures, in contrast to the results for IκBα. In this case, it was not necessary to include the rest of the dimerization domain of p65. The structures contained nearly all of the contacts found in the crystal structure (Figure 6).[27] In 90% of the cases, the p65 NLS polypeptide assumes the crystal structure-like conformation corresponding to the strong electron density. The RMSD of the cluster center from the crystal structure considering the NLS residues alone was 1.7 Å. When the interface

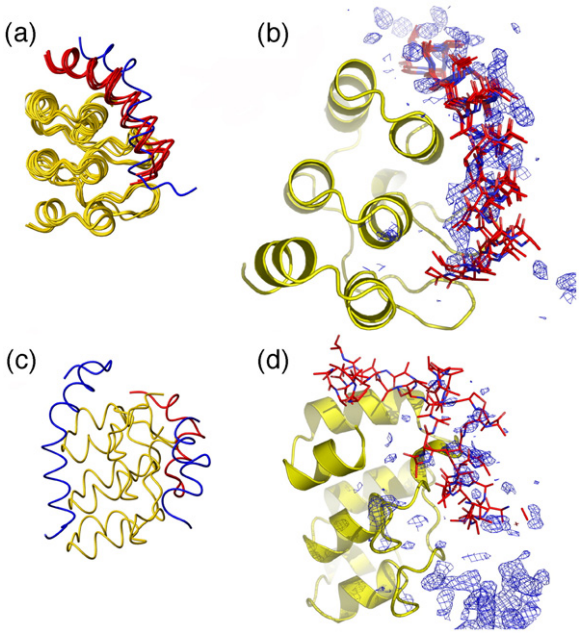

Figure 6. The phylogenetic tree was obtained using the Fitch-Margoliash distance-based algorithm for structures obtained in AMH simulations of the p65 NLS polypeptide bound to IκBβ. The tree shows clustering into one main group and one less prevalent alternative binding mode. (a) The dominant cluster of structures was obtained for the interaction between the p65 NLS polypeptide (red) and IκBβ (yellow). For comparison, the structure of the p65 NLS polypeptide structure from the NF-κB (p65/p65) IκBβ complex is shown (blue).[27] (b) Electron density map (blue) of the NLS polypeptide for $2F_o$–F_c density using model-derived phases for 1oy3 (PDB). IκBβ is shown in cartoon style in yellow, while the p65 NLS polypeptides obtained in AMH simulations are shown as red sticks. (c) The alternative symmetric binding site of the NLS polypeptide was sampled and compared reasonably well to (d) the existing electron density derived from the crystal structure.

residues were also included, the RMSD was 2.1 Å. A few (8%) of the cases sampled alternative structures comparable to the cluster 3 structures seen in the simulations of the p65 NLS with IκBα and similar to the minor electron density in the crystal structure. The RMSD from the crystal structure model of this weaker binding site to the cluster center for this alternative binding mode was 4.6 Å for the NLS only, and 5.4 Å for the NLS and interface residues (Figure 6(d)). For comparison, the RMSD from the alternative binding site of the basin 3 structure was 4.6 Å and 6.1 Å, respectively, for NLS only and NLS with interface residues. The structure of this alternative binding site was very similar to the cluster 3 binding mode, yielding a 1.2 Å RMSD from each other, and 2.8 Å when comparing the deviation of the NLS and the interface residues.

Binding of the p50 NLS polypeptide to IκBα/β

Since NF-κB dimers can be homo- or heterodimers composed of both p65 and p50, we performed AMH simulations of the folding and binding of the p50 NLS polypeptide to IκBα. Simulations of the p50 NLS tethered to just IkBa gave two main clusters of structures; 32% of the structures were similar to basin 4 in the p65 NLS simulations, and the remainder were similar to basin 2 structures. When the p50 dimerization domain was included to geometrically limit the search space of the p50 NLS, some 75% of these failed to adopt stable strong interactions between the p50 NLS polypeptide and IκBα whereas the presumed "native" structure was found in the remaining 25% of the structures. Simulations of the p50 NLS polypeptide interacting with IκBβ gave no preferred structures, with less than 25% of the structures falling into any one basin. The results suggest that the p50 NLS polypeptide can bind to IκBα with a structure similar to that observed for the p65 NLS bound to IκBα but the energy landscape is not as funneled as for the p65 NLS.

Discussion

The p65 NLS polypeptide has a high propensity to form helical structure

Folding simulations performed with the AMH on the p65 NLS polypeptide by itself showed that the p65 NLS polypeptide has a significant propensity to fold into a helix. The p50 NLS polypeptide also had helical propensity but did not contain the kink required for correct binding to IκBα/β. The structural role of the KRKR NLS sequence in p65 must break up this helical structure into two helices by forming a kink region. The p50 NLS contains only three basic residues flanked by Gln, and this slightly different sequence does not appear to be sufficient to break the helical structure. Binding of the polypeptide to the inhibitor requires the polypeptide to take on a fold of two helices connected by a break region. The two helices of the NLS polypeptide each bind to IκB, where the first helix forms stable contacts with the ankyrin repeats, while the second helix caps the hydrophobic top of the inhibitor. Our simulations show one role of the NLS sequence is to provide a break in the helical secondary structure for its interaction with IκB and help to explain the weaker binding affinity of the p50 NLS polypeptide.[10] The nucleoplasmin NLS polypeptide did not form one unique structure in the AMH simulations, and showed mostly extended structure consistent with the extended structure it adopts when bound to importins.[14] These results beg the question of whether the NF-κB NLS polypeptides bind to importins in an extended structure and therefore must be unraveled for binding, or whether they bind in a helical conformation.

Use of AMH to predict binding conformations

An important aspect of the present computer simulation study of binding and folding is that

reasonable predictions of the correct binding/docking site rely on the calibration of the energy function as reliable by examining other structurally well defined binding situations. Blind trials performed in CASP have already allowed us to have an unbiased evaluation of the of the *ab initio* AMH energy function. These comparisons have confirmed the predictive power of the AMH to obtain the folded state of largely helical proteins up to ~180 residues with high fidelity (M.C.P. and P.G.W., unpublished results). By constraining the problem using glycine linkers and keeping the binding partner fixed, we show here that this same energy function has predictive power for simple protein–protein interactions that involve folding as well as binding of surface helices.

Although the AMH energy function is not generally useful for inferring binding propensities, it is interesting to note that the folding of NF-κB NLS polypeptides upon binding to IκBβ could be compared directly to electron density maps for the IκBβ complex with NF-κB(p65/p65) (PDB 1oy3) in which two binding sites of the p65 NLS polypeptide are observed, one with strong electron density (Figure 6(a) and (b)) and a second with weak electron density (Figure 6(c) and (d)). The simulations are consistent with the electron density results, with some 90% of the structures binding to the site with the stronger electron density. while the rest of the predicted structures closely resemble the binding site with weaker electron density.

Prediction of a second p65 NLS polypeptide binding site on IκBα

Ernst *et al.* showed that one molecule of IκBα binds to one NF-κB dimer containing two NLS polypeptides.[28] Our simulations of the p65 NLS binding to IκBα find a major cluster that corresponds to the conformation found in the IκBα-NF-κB(p50/p65) crystal structure and a second symmetric binding site similar to that seen in the binding to IκBβ that was observed crystallographically. The folding of NF-κB NLS polypeptides upon binding to IκBβ could be compared directly to the crystal structure of the IκBβ-NF-κB(p65/p65) homodimer.[27] The p65 NLS polypeptide binds strongly to IκBβ in the crystal structure-like conformation (Figure 6). Comparison of electron density maps for the IκBβ-NF-κB(p65/p65) (1oy3.pdb) with the structures obtained in our simulations demonstrate the power of the AMH to predict the location of both the strong (Figure 6(a) and (b)) and weak (Figure 6(c) and (d)) binding sites. The simulations show that the landscape of binding of the p65 polypeptide to IκBβ is strongly funneled, with some 90% of the structures binding to the site with the stronger electron density.

Thus, a new prediction made from this work is that when the X-ray crystal structure of IκBα with NF-κB (p65/p65) is completed, there will be two NLS binding sites observed on IκBα as were observed on IκBβ. This prediction awaits confirmation when the crystal structure is determined, but is consistent with the speculation that IκBα might mask both NLS polypeptides, resulting in more complete cytoplasmic localization of the p65/p65 homodimer.[29]

In contrast to the p65 NLS polypeptide, the p50 NLS polypeptide interaction energy landscape was less funneled. These results are consistent with the observation that the p50 NLS polypeptide does not contribute to the binding energy of the NF-κB(p50/p65)/IκBα complex,[10] and that the p50 NLS remains unbound and is responsible for the NLS-dependent import into the nucleus, resulting in continuous shuttling of the NF-κB(p50/p65)/IκBα complex.[17,29,30] In some simulations, the p50 NLS polypeptide interacted with the "native-like" binding site on IκBα, and this result is consistent with the report that the p50 NLS is required for NF-κB(p50/p50) binding to IκBα.[31] This prediction may help resolve a controversy in the NF-κB field because previous biochemical data have given the impression that IκBα and IκBβ are distinct largely because of how many NLS sequences they could sequester. In contrast, experiments in transgenic mice indicated that the two proteins were functionally equivalent.[32] Our results predict that they are functionally equivalent, even in the number of NLS sequences that they sequester, and that the biochemistry was misleading due to an incomplete set of dimers tested.

The p65 NLS polypeptide finds its correct binding site on IκBα/β in the absence of the rest of the p65 molecule

We have not examined explicitly how IκB and DNA compete for binding to NF-κB. The simulations suggest that the NF-κB subunit NLS polypeptide, independent of the rest of the NF-κB molecule, is capable of spontaneously undergoing a disordered to ordered transition upon binding to the inhibitor. The simulations further predict that it is the p65 subunit NLS polypeptide that conveys the binding specificity observed between the IκBα and NF-κB subunits. It is possible that even when the NF-κB is bound to DNA, the unstructured p65 NLS polypeptide may interact with IκBα, ultimately facilitating the disassembly of the enhanceosome complex. The IκB/NF-κB system represents a beautiful example of how induced fit is used by Nature to achieve control of binding specificity through structural diversity.

Materials and Methods

Protein constructs and sequences

Simulations of the dynamics of several different constructs were performed using the associative memory hamiltonian (AMH).[16] These constructs included the 30 residue NLS polypeptide of the p65 subunit (residues 291–320, chain C, PDB 1NFI) of the p65–p50 complex (1nfi,[12] by itself, and when linked to the first three ankyrin repeats (residues 70–156, chain D, PDB 1NFI) of the ankyrin repeat domain of IκBα *via* a glycine linker. A truncated p65

fragment containing the NLS linked *via* an N-terminal glycine linker to the C terminus of the truncated IκBα fragment was also simulated, as well as a construct where the NLS polypeptide was replaced by the nucleoplasmin NLS. The initial configurations of these constructs were all built using the Biopolymer module of Insight II†.

For the first construct, a simple glycine chain connects the C-terminal Lys320 of the NLS polypeptide to the N-terminal Ser70. The advantage of using glycine residues is that they are much less sterically hindered than any other amino acid residue. The length and flexibility of the glycine chain allowed the NLS polypeptide to bind geometrically at any possible location on the IκBα protein surface. In later simulations, residues 192–320 of p65 and IκBα residues 70–156 were connected *via* a glycine linker connecting residue 156 of the C terminus of IκBα to residue 192 of the N terminus of p65. The purpose of this construct was to further investigate trap states in the folding and binding of the NLS polypeptide that might not be available when IκBα and p65 were arranged in crystal structure-like geometry. The same constructs with the p65 NLS polypeptide exchanged for the NLS polypeptide of p50 and for nucleoplasmin (PDB 1EE5) were also simulated.

Simulated annealing protocols with the associative memory Hamiltonian

The associative memory hamiltonian (AMH) is an energy function designed for making *ab initio* predictions of 3D protein structure from a given amino acid sequence. The AMH is an optimized energy function used for protein structure prediction even in the absence of homology information.[16] The terms of the full energy function used for the simulation contain besides the AMH sequence dependent interaction, V_{AM} for short-range and medium-range interactions, and V_{contact} for long-range contact interactions, and excluded volume terms V_{ev} and basic backbone terms that include a potential $V_{\phi\varphi}$, which provides a good fit of the backbone torsion angles found in a Ramachandran map, and hydrogen bonding patterns to assure correct physics and chemistry of the polypeptide chain. Here, the excluded volume potential is applied to the carbon and oxygen atoms that approach within 3.5 for $(j–i)<5$, and 4.5 for $(j–i)\geq 5$. The chirality potential V_{χ} biases the peptide chain into the L-amino acid configuration. V_{harmonic} contains three quadratic potentials along with shake constraints for the heavy backbone atoms to provide backbone rigidity. The total potential used for the AMH molecular dynamics simulations is given by:

$$V_T = V_{\mathrm{AM}} + V_{\mathrm{contact}} + \lambda_{\phi\phi}V_{\phi\phi} + \lambda_{\chi}V_{\chi} + \lambda_{\mathrm{ex}}V_{\mathrm{ex}} + \lambda_{\mathrm{harmonic}}V_{\mathrm{harmonic}}$$

The λ -terms scale the strength of interaction of the individual potential terms. The functional form of the terms in the potential has been described.[16] The AMH potential uses different interactions for pairs separated by different numbers of residues in the sequence. The short-range and medium-range potential applies to residue pairs less than 12 residue apart in the sequence. These predict formation of local secondary structure such as helices and turns. Residues that are more than 12 residues apart in the sequence interact *via* contact interactions that contribute to the collapse of the protein and form tertiary structures from the shorter units. The equation for the associative memory term is:

$$V_{\mathrm{AM}} = -\sum_{\mu}^{n}\sum_{i<j}^{N}\gamma(P_i,P_j,P_{i'}^{\mu},P_{j'}^{\mu})\Theta(r_{ij}-r_{i'j'}^{\mu})$$

where μ runs over n memories. The parameters γ are learned by an optimization procedure.[16] The parameters are functions of P, where P represents the four-letter code designation assigned to each of the 20 naturally occurring amino acids.[33] The specific amino acids in each category are hydrophilic (Ala, Gly, Pro, Ser, and Thr), hydrophobic (Cys, Ile, Leu, Met, Phe, Trp, Tyr, and Val), acidic (Asn, Asp, Gln, and Glu), and basic (Arg, His, and Lys). The VAM potential encodes these sequence patterns by measuring the structural similarity to a list of memory proteins. This similarity is expressed functionally in the Θ function, which is a centered Gaussian that depends on the difference of distances in the simulated protein structure from those found in the memory protein. The contributions of the contact term to the total potential is given by a three-well potential:

$$V_{\mathrm{contact}} = -\frac{\varepsilon}{a}\sum_{i<j-12}\sum_{k=1}^{3}\gamma(P_i,P_j,k)c_k(N)U\left[r_{\min}(k),r_{\max}(k),r_{ij}\right]$$

These interactions are weighted by γ, depending on spatial distance and amino acid interaction type. The parameters in the potential are optimized using the quantitative form of the principle of minimal frustration, to yield the most funnel-like landscape for folding as possible, while maintaining transferability from one sequence to another. The details of the parameters of the potential have been described.[16] U is a contact function that controls the sharpness of the $k=3$ well potential at the potential boundary endpoints $r_{\min,\max}(k)$. The $c_k(N)$ terms are found from fitting the number of contacts in each of the regions as a function of sequence length of the target protein.

The physical principles of energy landscape theory apply to folding and to binding processes. Here, the AMH energy function constructed originally for folding prediction is applied to a two-protein construct in which the binding partners are fused with a variable glycine linker. The results document that the AMH predicts the correct crystal structure of binding, and does allow some possible alternative, thermodynamically plausible, binding modes. The AMH short-range and medium-range interactions fold the local helices of the NLS polypeptide correctly, while the long-range potential contact interactions of the AMH can dock and bind the p65 NLS polypeptide to the surface of IκBα correctly.

The simulation protocols were as follows: initially, 130 annealing runs of the NLS/IκBα constructs were performed. An additional 150 annealing runs with linker lengths of five and 13 glycine residues, were performed to test for any dependence of the results on linker length. To investigate the traps, 60 additional annealing runs were carried out for all the other constructs. Each individual annealing run trajectory sampled 280 independent structures. This resulted in a total of over 100,000 structures available for analysis. In the annealing runs of the complexes, the NLS polypeptide was initially unfolded randomly and placed as far away from IκBα as possible. Initial random velocities were assigned to the protein. Temperature is quoted in units of the native state energy per residue obtained as an average over the memory terms

† http://www.accelrys.com/

containing only associative memory terms, which are short-range and medium-range in sequence, and the contact potential. Defining the scaled quantity ε, the native state energy of all the memory energy terms including contact terms, for a protein of N residues as:

$$\varepsilon = \left(\frac{\overline{E^{\text{Native}}_{\text{AM+C}}}}{4N}\right)$$

then a reduced temperature T^* can be defined with $k_B T = \varepsilon T^*$.[34] All other energy terms, such as the backbone terms and excluded volume terms, are scaled to yield physically reasonable interaction strengths. The temperature T^*, where $T^* \sim 1$ is of the order of the folding temperature, was then reduced linearly from $T^* = 1.7$ to $T^* = 0.0$, resulting in trajectories of several hundred μs. This timescale was sufficient for both binding and folding of the NLS polypeptide to occur. A constraining potential assured that the three ankyrin repeats of IκBα/β did not change the topology of the backbone C^α atoms during the molecular dynamics annealing runs. The motivation for this constraint is based on the experimental fact that ankyrin repeats 2 and 3 of free IκBα show less hydrogen/deuterium exchange and thus higher protection factors when compared to the rest of the protein, suggesting an at least partially folded character of this region of the ankyrin repeat domain.[21] The side-chains of even the constrained part of the protein were free to move and interact with the flexible binding partner, which allows important interactions between the side-chains of the binding partners to occur.

Topological comparison

The overall topology of the protein fold and the secondary and tertiary structure of a protein can be monitored quantitatively by a variety of means. The RMSD is a standard way to probe the deviation between two structures but, since it is based on Gaussian statistics, it is best for very close structures. The fraction of overlapping structured pairs in two different structures is used as another measure of similarity, which is energetically relevant because the interactions are dominantly pairwise terms.

A useful, normalized quantity is Q, which is 1 for two identical structures and 0 for two structures that have no pair distances in common.[33] It is defined by:

$$Q = \frac{2}{(N-1)(N-2)} \sum_{i<j-1} \exp\left(-\frac{(r_{ij} - r_{ij}^N)^2}{2\sigma_{ij}^2}\right)$$

Here, the sum is over residue pairs that are separated by at least two residue in the sequence, with r_{ij} being the C^α distance of residues i and j, N is the total number of residues in the target structure and σ_{ij} is the Gaussian variance and is defined as (given in Å):

$$\sigma_{ij} = |i-j|^{0.15}$$

The CE Z-score obtained from a combinatorial extension (CE) algorithm is a statistical measure of how similar are the topologies and secondary structure of two proteins; a Z-score of 3.5–4 indicates possible biologically interesting similarities, naturally occurring proteins with structural Z-score greater than 4.0 are usually part of a single protein family or superfamily, and share the same overall topology and secondary structure.[35]

Structural clustering analysis

The Fitch-Margoliash algorithm is a distance-based bioinformatics algorithm to fit a phylogenetic tree to a distance matrix.[36] The numerous structures obtained from the annealing runs were clustered using the Fitch program of the PHYLIP package‡. The Fitch program is an algorithm that is designed to create phylogenetic trees based on a distance measure. In order to analyze the structures obtained in the simulated annealing with the bioinformatics software, a distance measure d between two structures A and B was introduced as $d = 1 - Q$. Since Q is a normalized measure of the fraction of overlapping contacts, d is a measure of how dissimilar two structures are, which is equivalent to a distance measure, if distance is taken in the sense that similar structures with small d are close and dissimilar structures with large d are far away. In the calculation of Q, the contacts mediated by glycine residues were excluded. The clustering was carried out using different choices of Q. For example Q_{total} takes into account all contacts, while $Q_{\text{interface}}$ includes only the interfacial contacts between the nuclear localization signal and IκBα. Eastwood *et al*. point out that flexible fragments cause little change in the number of native contacts, but give rise to large fluctuations in RMSD.[34]

Free energy calculations

Free energy profiles were obtained using the weighted histogram analysis method WHAM, which is a combination of free energy perturbation and umbrella sampling, and allows accurate and efficient calculation of the free energy profile from a given set of simulations. The method was as follows: 17 constant-temperature molecular dynamics simulations were performed with a polynomial Q biasing potential of fourth order centered on different values of Q ($Q = 0.1, 0.15, 0.2, \ldots, 0.9$) to obtain good phase-space sampling along this reaction coordinate. During each simulation, 200 samples, N_i^{obs}, of Q and energy E, the backbone and AMH energy, were collected at regularly spaced time-points. The first 40 samples were discarded to allow for equilibration. A histogram $N_i(E,Q)$ for all 17 simulations was created, which gave the density of states $n(E,Q)$ of the system:[34]

$$n(E,Q) = \sum_i w_i(E,Q) \frac{N_i(E,Q)}{N_i^{\text{obs}}} e^{\beta_i E + \beta_i V_i(Q)} Z_i(\beta_i)$$

with i being the index of simulation. The partition function is given by:

$$Z_i = \sum_{E,Q} n(E,Q) e^{-\beta_i E - \beta_i V_i(Q)}$$

This allowed for self-consistent determination of the density of states $n(E,Q)$ to within a multiplicative constant and, hence, the free energy was obtained to within a constant as:

$$F(Q,T) = -k_B T \log\left(\sum_{E,Q} n(E,Q) e^{-\frac{E}{k_B T}}\right)$$

The free energy function thus obtained is now a function of the desired order parameter Q and temperature T.

‡ http://evolution.genetics.washington.edu/phylip.html

Given a temperature, the free energy profile allows the identification of thermodynamically distinct states of the protein, such as the unfolded state ensemble.

B-factor calculations

The **B**-factor can be related to the mean displacement of a structure obtained in molecular dynamics simulations by the Debye-Waller equation:

$$\boldsymbol{B} = \frac{8}{3}\pi^2\langle R^2\rangle$$

where $\langle R^2\rangle = \langle (R_i - R_0)^2\rangle$ is the mean fluctuation in distance of the backbone atoms of structures sampled with the AMH relative to the refined crystal structure R_0.

Electron density maps

Reflection data of the PDB file 1OY3 was downloaded from the Protein Data Bank and converted into $2F_o$–F_c electron density maps with the program CNS.[37] Electron maps were drawn with the PyMOL software package§.

Acknowledgements

We thank Gourisankar Ghosh, Tom Huxford and Alexander Hoffmann for helpful discussions and critical reading of the manuscript. We thank De Bin Huang for help with creating the electron densities of the bound NLS. Support from the La Jolla Interfaces in Science (LJIS) and the R01 GM44557 NIH grant is gratefully acknowledged. J.L. thanks the National Science Foundation-sponsored Center for Theoretical Biological Physics grants PHY-0216576 and 0225630, and the CTBP for computer time on their cluster.

References

1. Dang, C. V. & Lee, W. M. F. (1988). Identification of the human c-Myc protein nuclear translocation signal. *Mol. Cell. Biol.* **8**, 4048–4054.
2. Dingwall, C. & Laskey, R. A. (1991). Nuclear targeting sequences—a consensus. *Trends Biochem. Sci.* **16**, 478–481.
3. Kalderon, D., Richardson, W. D., Markham, A. F. & Smith, A. E. (1984). Sequence requirements for nuclear location of simian virus-40 large-T-antigen. *Nature*, **311**, 33–38.
4. Lanford, R. E. & Butel, J. S. (1984). Construction and characterization of an SV40 mutant defective in nuclear transport of T-antigen. *Cell*, **37**, 801–813.
5. Leung, S. W., Harreman, M. T., Hodel, M. R., Hodel, A. E. & Corbett, A. H. (2003). Dissection of the karyopherin alpha nuclear localization signal (NLS)-binding groove–functional requirements for NLS binding. *J. Biol. Chem.* **278**, 41947–41953.
6. Baeuerle, P. A. & Baltimore, D. (1988). I-Kappa-B–a specific inhibitor of the Nf-Kappa-B transcription factor. *Science*, **242**, 540–546.
7. Baeuerle, P. A. & Baltimore, D. (1988). Activation of DNA-binding activity in an apparently cytoplasmic precursor of the Nf-Kappa-B transcription factor. *Cell*, **53**, 211–217.
8. Huxford, T., Malek, S. & Ghosh, G. (1999). Structure and mechanism in NF-kappa B/I kappa B signaling. *Cold Spring Harbor Symp. Quant. Biol.* **64**, 533–540.
9. Malek, S., Huxford, T. & Ghosh, G. (1998). IkBa functions through direct contacts with the nuclear localization signals and the DNA binding sequences of NF-kB. *J. Biol. Chem.* **273**, 25427–25435.
10. Phelps, C. B., Sengchanthalangsy, L. L., Huxford, T. & Ghosh, G. (2000). Mechanism of I kappa B alpha binding to NF-kappa B dimers. *J. Biol. Chem.* **275**, 29840–29846.
11. Huang, D. B., Huxford, T., Chen, Y. Q. & Ghosh, G. (1997). The role of DNA in the mechanism of NF kappa B dimer formation: crystal structures of the dimerization domains of the p50 and p65 subunits. *Structure*, **5**, 1427–1436.
12. Jacobs, M. D. & Harrison, S. C. (1998). Structure of an I kappa B alpha/NF-kappa B complex. *Cell*, **95**, 749–758.
13. Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54–60.
14. Conti, E. & Kuriyan, J. (2000). Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. *Struct. Fold. Des.* **8**, 329–338.
15. Huxford, T., Huang, D. B., Malek, S. & Ghosh, G. (1998). The crystal structure of the I kappa B alpha/NF-kappa B complex reveals mechanisms of NF-kappa B inactivation. *Cell*, **95**, 759–770.
16. Hardin, C., Eastwood, M. P., Luthey-Schulten, Z. & Wolynes, P. G. (2000). Associative memory Hamiltonians for structure prediction without homology: alpha-helical proteins. *Proc. Natl Acad. Sci. USA*, **97**, 14235–14240.
17. Malek, S., Chen, Y., Huxford, T. & Ghosh, G. (2001). I kappa B beta but not I kappa B alpha, functions as a classical cytoplasmic inhibitor of NF-kappa B dimers by masking both NF-kappa B nuclear localization sequences in resting cells. *J. Biol. Chem.* **276**, 45225–45235.
18. Koshland, D. E. (1995). The key-lock theory and the induced fit theory. *Angew. Chem.* **33**, 2375–2378.
19. Papoian, G. A. & Wolynes, P. G. (2003). The physics and bioinformatics of binding and folding–an energy landscape perspective. *Biopolymers*, **68**, 333–349.
20. Levy, Y., Wolynes, P. G. & Onuchic, J. N. (2004). Protein topology determines binding mechanism. *Proc. Natl Acad. Sci. USA*, **101**, 511–516.
21. Croy, C. H., Bergqvist, S., Huxford, T., Ghosh, G. & Komives, E. A. (2004). Biophysical characterization of the free IkappaBalpha ankyrin repeat domain in solution. *Protein Sci.* **13**, 1767–1777.
22. Bergqvist, S., Hughes, C., Huxford, T., Ghosh, G. & Komives, E. (2004). Thermodynamics and kinetics of the NF-kB/IkB and NF-kB/DNA interactions. *Biophys. J.* **86**, 513a.
23. Kim, D., Xu, D., Guo, J., Ellrott, K. & Xu, Y. (2003). PROSPECT II: protein structure prediction program for genome-scale applications. *Protein Eng.* **16**, 641–650.
24. Liou, H. C., Nolan, G. P., Ghosh, S., Fujita, T. & Baltimore, D. (1992). The NF-kappa B p50 precursor, p105, contains an internal I kappa B-like inhibitor that preferentially inhibits p50. *EMBO J.* **11**, 3003–3009.
25. Janin, J. (2002). Welcome to CAPRI: a critical

§ http://pymol.sourceforge.net/

assessment of PRedicted interactions. *Proteins: Struct. Funct. Genet.* **47**, 257.
26. Cheng, J. D., Ryseck, R. P., Attar, R. M., Dambach, D. & Bravo, R. (1998). Functional redundancy of the nuclear factor kappa B inhibitors I kappa B alpha and I kappa B beta. *J. Expt. Med.* **188**, 1055–1062.
27. Malek, S., Huang, D. B., Huxford, T., Ghosh, S. & Ghosh, G. (2003). X-ray crystal structure of an I kappa B beta center dot NF-kappa B p65 homodimer complex. *J. Biol. Chem.* **278**, 23094–23100.
28. Ernst, M. K., Dunn, L. L. & Rice, N. R. (1995). The pest-like sequence of I-Kappa-B-Alpha is responsible for inhibition of DNA-binding but not for cytoplasmic retention of C-rel or rela homodimers. *Mol. Cell. Biol.* **15**, 872–882.
29. Tam, W. F., Wang, W. & Sen, R. (2001). Cell-specific association and shuttling of IkappaBalpha provides a mechanism for nuclear NF-kappaB in B lymphocytes. *Mol. Cell. Biol.* **21**, 4837–4846.
30. Johnson, C., Van Antwerp, D. & Hope, T. J. (1999). An N-terminal nuclear export signal is required for the nucleocytoplasmic shuttling of IkappaBalpha. *EMBO J.* **18**, 6682–6693.
31. Latimer, M., Ernst, M. K., Dunn, L. L., Drutskaya, M. & Rice, N. R. (1998). The N-terminal domain of I kappa B alpha masks the nuclear localization signal(s) of p50 and c-Rel homodimers. *Mol. Cell. Biol.* **18**, 2640–2649.
32. Cheng, J. D., Ryseck, R. P., Attar, R. M., Dambach, D. & Bravo, R. (1998). Functional redundancy of the nuclear factor kappa B inhibitors I kappa B alpha and I kappa B beta. *J. Expt. Med.* **188**, 1055–1062.
33. Cho, S. S., Levy, Y. & Wolynes, P. G. (2006). P versus Q: structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl Acad. Sci. USA*, **103**, 586–591.
34. Eastwood, M. P., Hardin, C., Luthey-Schulten, Z. & Wolynes, P. G. (2001). Evaluating protein structure-prediction schemes using energy landscape theory. *IBM J. Res. Develop.* **45**, 475–497.
35. Shindyalov, I. N. & Bourne, P. E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**, 739–747.
36. Fitch, W. M. & Margoliash, E. (1967). Construction of phylogenetic trees. *Science*, **155**, 279–284.
37. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (2005). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.

Edited by B. Honig

(Received 8 March 2006; received in revised form 28 November 2006; accepted 4 December 2006)
Available online 15 December 2006