Published in final edited form as:

J Am Chem Soc. 2017 December 27; 139(51): 18558–18566. doi:10.1021/jacs.7b08741.

Resolving the NF κ B Heterodimer Binding Paradox: Strain and Frustration Guide the Binding of Dimeric Transcription Factors

Davit A. Potoyan^{†,‡,iD}, Carlos Bueno[‡], Weihua Zheng[‡], Elizabeth A. Komives[§], and Peter G. Wolynes^{*,‡}

[†]Department of Chemistry, Iowa State University, Ames, Iowa 50011, United States

[‡]Department of Chemistry and Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States

[§]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, California 92093, United States

Abstract

Many eukaryotic transcription factors function after forming oligomers. The choice of protein partners is a nonrandom event that has distinct functional consequences for gene regulation. In the present work we examine three dimers of transcription factors in the NF κ B family: p50p50, p50p65, and p65p65. The NF κ B dimers bind to a myriad of genomic sites and switch the targeted genes on or off with precision. The p65p50 heterodimer of NF κ B is the strongest DNA binder, and its unbinding is controlled kinetically by molecular stripping from the DNA induced by I κ B. In contrast, the homodimeric forms of NF κ B, p50p50 and p65p65, bind DNA with significantly less affinity, which places the DNA residence of the homodimers under thermodynamic rather than kinetic control. It seems paradoxical that the heterodimer should bind more strongly than either of the symmetric homodimers since DNA is a nearly symmetric target. Using a variety of energy landscape analysis tools, here we uncover the features in the molecular architecture of NF κ B

*Corresponding Author. pwolynes@rice.edu.

ORCID

Davit A. Potoyan: 0000-0002-5860-1699

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b08741.

Simulation protocols, description of AWSEM+3SPN.2C force fields, and additional figures and references (PDF)

Heterodimer-Strain-PC1 (AVI)

Heterodimer-Strain-PC2 (AVI)

p50p50-frustration-CartesianPC1 (AVI)

p50p50-frustration-CartesianPC2 (AVI)

p50p65-frustration-CartesianPC1 (AVI)

p50p65-frustration-CartesianPC2 (AVI)

p65p65-frustration-CartesianPC1 (AVI)

p65p65-frustration-CartesianPC2 (AVI)

The authors declare no competing financial interest.

dimers that are responsible for these drastically different binding free energies. We show that frustration in the heterodimer interface gives the heterodimer greater conformational plasticity, allowing the heterodimer to better accommodate the DNA. We also show how the elastic energy and mechanical strain in NF- κ B dimers can be found by extracting the principal components of the fluctuations in Cartesian coordinates as well as fluctuations in the space of physical contacts, which are sampled via simulations with a predictive energy landscape Hamiltonian. These energetic contributions determine the specific detailed mechanisms of binding and stripping for both homo- and heterodimers.

Graphical abstract

INTRODUCTION

Transcription factor oligomers regulate the flow of genetic information.¹ Even at the dimer level, different combinations of partners bind differently to specific genomic regions and can thereby trigger different sequences of downstream events, eventually leading to different cell fates.^{1,2} There are many reasons why forming and acting as oligomers may be biologically adaptive for transcriptional regulation. The combinatorial diversity of pairing can allow a finer tuning of gene expression by turning a limited number of transcription factor coding genes into a much larger pool of disparate transcription factor complexes with widely varying binding specificity toward different regulatory genomic sequences. The association of proteins by itself can then serve as a switch or a control point for exerting further levels of post-translational regulation.² Finally, oligomerization can be seen as adding nonlinearity to the system-wide kinetic network, which allows the response to changes in protein copy numbers to be more sensitive.³ Such nonlinearity also offers noise buffering benefits^{4,5} and can turn steady states of regulatory networks into oscillatory attractor patterns.⁶ We thus see that uncovering the fundamental principles behind transcription factor association and DNA recognition may lead to a deeper understanding of basic cell biology and perhaps may allow one to devise better therapeutic strategies to intervene when regulatory networks fail.

In this Article, we take a look at these basic questions by focusing on one specific important gene regulatory system that involves the transcription factor NF κ B. This transcription factor is the central regulator of an extensive and complex genetic broadcasting system.⁷ We have shown that the behavior of this broadcasting network is controlled by the kinetic features of transcription factor binding and release from DNA.^{8,9} The NF κ B family of proteins has many members, and these function in various combinations of homodimers and heterodimers. The heterodimeric form, made of two different subunits, p50p65, happens to be the most abundant in cells. This heterodimer binds to DNA significantly more strongly than do either of the p50p50 and p65p65 homodimers.¹⁰ The steady-state ratio of different dimers in cells is clearly controlled via intricate gene regulatory pathways, which cannot be explained solely by DNA binding affinities or dimerization equilibrium constants.¹¹ *A priori*, if binding to DNA were made up of individual additive contributions from each monomer, we would expect the heterodimer to have an affinity somewhat between the affinities of the two homodimers, not to be dramatically stronger than either of them. The fact that the heterodimer binds more strongly than either homodimer is, therefore, something of a paradox.

Using a combination of gel mobility shift and fluorescence anisotropy assays, Phelps et al.^{10,12} have established that p65p50 binds to the κ B target site of immunoglobulin enhancer with an affinity of ~10 nM; p50p50 has a weaker affinity of ~50 nM, and p65p65 has the weakest affinity of all three dimers at ~150 nM. These measurements were all done at near-physiological conditions. The genome-wide patterns of binding are expected to be more complex, with different dimers displaying a distribution of affinities that show mutually exclusive preferences for some sequences. Nevertheless, protein binding microarray experiments¹³ have shown that the distributions for 10-bp κ B sites with different flanking bases still place p65p50 consistently as a stronger binder than both the p50p50 and p65p65 dimers. The tight binding of the heterodimeric NF κ B complex is apparently not an accident. In fact, cells have elaborated a mechanism we have termed “molecular stripping” to rapidly and completely remove the heterodimeric NF κ B from its many genomic target sites^{8,9,14,15} (Figure 1A). The relative concentration of free and bound heterodimeric forms is thus under kinetic control. The homodimers, on the other hand, because of their weak binding, do not need molecular stripping, since their dissociation is sufficiently rapid¹⁶ (Figure 1A). Their relative concentrations of DNA-bound and free forms can equilibrate thermodynamically. The steady-state composition of the different dimers will thus be seen to be under kinetic, not thermodynamic, control owing to the strong heterodimeric association and molecular stripping.

In the present work, we are concerned with understanding the basic molecular features that resolve the “binding paradox” by using predictive coarse-grained models for energy landscapes of the proteins and the DNA in order to simulate the structural and mechanical changes that accompany the dissociation of these transcription factors from the DNA-bound complexes.

To set the stage for describing the nature of these changes and the computational tools used for analyzing them, it is instructive to first briefly survey some basic tenets of the theory of protein dynamics. The modern view of protein dynamics is provided by energy landscape

theory.¹⁷ A key notion from this theory is that the landscape overall resembles a funnel, allowing rapid sampling of the global conformational space of foldable proteins. The funnel-like shape of the global energy landscape is a result of the evolutionary minimization of the conflict between different residual interactions in proteins, which is also known as “frustration”. This principle of minimal frustration^{18,19} explains that evolution has sculpted funnel-like landscapes for proteins in order to eliminate deep kinetic traps, which would otherwise lead to unreasonably long folding times. The funnel topography ensures rapid and reliable folding into functional states. Because of the funnel-like shape of the landscape, most conformational changes can be described either as arising from distortions of the native structure or as arising from the making and breaking of native interactions; completely non-native structures are rare as mechanistic intermediates (but do arise on occasion). The funnel-like shape of the folding landscape can also be exploited for deriving potentials for protein structure prediction using machine learning strategies.²⁰ The resulting optimized landscapes provide predictive force fields that can be used to simulate functional dynamics, as is done in the present work. The bottom of the funnel has low but nonzero entropy and contains highly structured conformations which can occupy multiple distinct structural basins. The multiplicity of these basins arises partly from the residual frustration.²¹ The frustration can be readily quantified by computational tools such as the Frustratometer.²² Residual frustration is not necessarily an accident of nature coming from some failure of evolutionary optimization but can signal an adaptation to facilitate ligand-induced functional allosteric changes of protein conformational states.¹⁹ Because of the multibasin character of most functional proteins, the dynamics of native proteins is anisotropic.^{21,23} While individual minima can be well described by quasiharmonic local fluctuations with a good fit to crystallographic *B* factors, jumps among the different basins are accomplished via larger amplitude changes also known as “proteinquakes”.^{24,25} Such large-amplitude motions involve high-order nonlinear deformations, “cracking”,^{26–29} and local unfolding,²¹ all of which require going beyond the harmonic approximations in the energy intrinsic to the idea of elasticity.^{21,23} These large-amplitude motions occur predominantly by moving along the directions of some of the few low-frequency soft modes which can be extracted by techniques such as principal component analysis (PCA) of fluctuations in long molecular dynamics simulation runs. Thus, one can use concepts from the elasticity theory of materials,²⁴ multibasin approximations,^{30,31} and linear response theory^{32,33} to predict and explain the basis of ligand-induced structural changes. Below we use an array of these tools to characterize the protein dynamics in the NF- κ B complexes to uncover the structural and dynamic resolution of the “binding paradox” of NF- κ B transcription factors. In particular, our analysis in some sense should also be considered an extension of the approach first introduced by Miyashita et al.,²⁴ where elasticity theory was employed for scrutinizing the allosteric movements of adenylate kinase in terms of mechanical strain and strain-induced cracking. A novelty of the present approach is the use of a nonstandard set of collective coordinates: the principal components of pair distances involved in the space of physical contacts. These new collective variables enable us to carry out very detailed quantification and visualization of how the mechanical strain results from stretching numerous individual physical contacts distributed within and between protein domains. Similarly, we introduce a binary-valued representation of the dynamical contacts (0 = broken, 1 = unbroken) in order to visualize the specific cracking transitions where numerous contacts become completely

broken. We call the principal components associated with the contact pair distances “strain PCs” and the principal components of the fluctuations of the binary contact observables “cracking PCs”. These two kinds of collective coordinates allow us to visualize the rich cooperative interplay between straining and cracking of specific contacts, shedding light on how the allosteric coupling between the dimer interaction interface and the DNA binding interface is made possible through soft modes of motion of different NF- κ B dimers, which then contribute to the differences in their binding affinities. We also pinpoint how energetic frustration in the heterodimeric interface allows the heterodimer to have greater flexibility at this key hinge, which enables the heterodimer to bind more intimately to its DNA targets than do the more rigid and unfrustrated homodimers, which cannot clamp down as effectively onto the DNA. We believe the tools introduced in the present analysis of the NF- κ B system shed light on the relationship between binding affinity and molecular architecture of transcription factors in general and also can be used to analyze and understand other systems. Our calculations can be further tested through protein engineering studies that probe how transcription factor–DNA interactions can be changed through seemingly remote mutations.

MATERIALS AND METHODS

We use physics-based, coarse-grained predictive protein and DNA models in order to simulate the process of DNA dissociation from the NF- κ B–DNA bound complexes as well as to generate equilibrium conformational ensembles for the endpoint states of the free and DNA-bound NF- κ B dimers. We use the predictive protein energy landscape model provided by the Associative memory, Water-mediated, Structure and Energy Model (AWSEM) developed in the Wolynes’s group³⁴ together with the 3 Sites Per Nucleotide (3SPN.2C) model for DNA developed by de Pablo’s group.^{35,36} The total Hamiltonian, $H_{\text{tot}} = H_{\text{AWSEM}} + H_{\text{3SPN.2C}} + H_{\text{pp-DNA}}$, consists of terms describing both the interactions within the protein components and within the DNA and the interactions between the two molecular species. Besides steric terms preventing molecular overlap, the interactions between protein and DNA are largely electrostatic. We have chosen to approximate these electrostatic interactions with a simple mean-field Debye–Hückel potential,

$$H_{\text{pp-DNA}} = \sum_{i,j} \frac{q_i q_j}{\epsilon r_{ij}} e^{-r_{ij}/\kappa_D}$$

where ϵ , the dielectric constant, is set to 80, and κ_D , the Debye length, is set to 10 nm in order to approximate the physical environment of the cell.

To carry out simulations, we use as initial input the crystal structures provided in the Protein Data Bank (PDB) for p65p65 (2RAM), p65p50 (1LE5), and p50p50 (1SVC) NF- κ B dimers. The raw PDB structures of the dimers, however, are missing important structural moieties which we had to computationally reconstruct before running the equilibrium sampling of dimers and their complexes. The reconstruction of structures with complete sequences has been made possible by using the predictive force field of AWSEM,³⁴ which allows us to rapidly locate and sample equilibrium configurations consistent with crystal structures.

For the constant-temperature simulations, we perform routine minimization and equilibration protocols for all structures before running long-time-scale production runs, which are then subjected to analysis. A 20-bp DNA fragment with the same sequence that was used in the *in vitro* experiments¹⁶ on molecular stripping is used for simulating DNA-bound NF- κ B complexes. The orientation of the DNA has been constrained, however, in all of the windows of the umbrella sampling simulations to remain parallel to the initial orientation in DNA-bound complex. By this constraining of the orientation, we exclude contributions from rotational entropy as the DNA dissociates, as well as finite chain effects of DNA, so as to make sure that one measures the free energy of DNA detachment without any contributions from sliding to the end of a short fragment.

We have carried out several kinds of PCA on the conformational ensembles generated via constant-temperature sampling (equivalent to 500 μ s of real time⁸) as well as on ensembles sampled along the DNA dissociation coordinate (equivalent to 10 μ s of real time⁸). Cartesian principal components are computed using the ordinary Cartesian coordinates of the C $_{\alpha}$ atoms of the long protein chains [$r^{3N}(t)$]. We then use the principal eigenvectors found for the free NF- κ B dimers as a basis for projecting the trajectories of complexes of NF- κ B bound with DNA. This allows us to compare directly how the same twist/breathing modes evolve upon binding to the DNA.

To compute the contact-based principal components, we have used the same structural ensembles as were used for the Cartesian PCA. First we use a trajectory-averaged structure to define the C $_{\alpha}$ Cartesian coordinates $\langle r^{3N} \rangle$. The coordinates of this averaged structure $\langle r^{3N} \rangle$ are then used to compute the set of all pairwise distances, d'_{ij} , which we use for fiducial purposes. Next these pairwise distances are filtered to exclude contacts with sequence spacing closer than four amino acid residues apart as well as those that are farther apart than 10–12 Å in the 3D averaged structure. Once these pairs of interactions are defined, the set of pairwise distances, filtered in the same way, is then computed for all the snapshots of the trajectory $d_{ij}(t)$. By this means, we have selected only those nonbonded contact pairs which contribute significantly to the energy of elastic strain. For the strain PCA, we monitor the absolute magnitudes of the contact distances $d_{ij}(t)$. To compute the cracking principal components, we use binary quantities $q_{ij} = 0, 1$ to indicate the extent of contact formation, where $q_{ij} = 0$ corresponds to a contact being broken when the pairs become too far separated, $d_{ij}(t) > 8$ Å, while $q_{ij} = 1$ corresponds to a contact having been formed with a close distance, $d_{ij}(t) < 8$ Å. Using these contact definitions, we can quantify the elastic energy of proteins by a uniform quasi-harmonic approximation,²⁴

$$\Delta E_{\text{elast}}(t) = \frac{1}{2} \left\langle \sum_{i,j} (d_{ij}(t) - d_{ij}^{\text{ref}})^2 \right\rangle$$

where $\langle \rangle$ denotes the ensemble average, and d_{ij}^{ref} denotes the pairwise distances of the reference structure. One can also estimate the fractional energetic contribution from any individual strain or cracking principal component, PC $_n(t)$, by using a similar quasi-harmonic approximation:

$$\Delta E_{PC_n}(t) = \frac{1}{2\lambda_n^2} \langle (PC_n(t) - PC_n^{\text{ref}})^2 \rangle$$

where PC^{ref} is the value for the average reference structure, and λ_n is the eigenvalue associated with the n th principal component.

RESULTS AND DISCUSSION

When binding to DNA or to other ligands, many proteins are forced to undergo large-scale conformational rearrangements, where some intramolecular contacts inevitably must break in order to form new contacts that are more suited to the ligand. This is the basis of allosteric communication.

The dissociation of stably bound protein–DNA complexes is a slow process. A rare activated event is needed for unbinding to occur. We therefore used umbrella sampling simulations with a center-of-mass distance restraint between the DNA and NF κ B dimers in order to extract relevant free energy profiles along this zeroth-order reaction coordinate (see Materials and Methods and Supporting Information). Through such simulations, we are able to determine binding affinities at least qualitatively. The most important result of this analysis is that simulations using the AWSEM force field actually do recapitulate the “paradoxical” pattern of dissociation observed in the experiments in which the heterodimer binds more strongly to the DNA than do either of homodimers (Figure 1B). The predicted differences in binding affinities mirror those found by experiment. The ratios of the predicted to the measured binding free energy,

$$\eta = \frac{\Delta F^{\text{AWSEM-3SPN.2C}}}{\Delta F^{\text{exp}}}$$

for the various dimers are as follows: The p65p50 heterodimer binds with a predicted affinity that is about right, $\eta \approx 1$. The predictions of the binding free energies for p50p50 and p65p65 are each larger than are the experimental values: for the p50p50 homodimer $\eta \approx 2$, while $\eta \approx 3$ for the p65p65 homodimer. We can see that, in quantitative terms, the magnitude of the paradoxical discrepancy is somewhat underestimated by the simulation in comparison with experiment, but nevertheless the predicted disparities in the binding energies show very clearly that the heterodimer binding is much stronger than binding of either homodimer, as is seen in the laboratory.

The model we use has no DNA sequence-specific effects other than those that would come from the sequence-dependent shapes of the DNA grooves that are encoded in the 3SPN.2C Hamiltonian. We note that NF κ B binding is known to be relatively promiscuous,³⁷ suggesting this is a reasonable approximation. While any additional specific interactions in the Hamiltonian could change the picture quantitatively, we see that the large free energy difference between the binding of the heterodimer and the homodimers is already captured by the present purely electrostatic sequence-independent model of the direct protein–DNA interactions (Figure 1B). These results then suggest that the model is sufficiently realistic to

justify extracting more of the relevant structural features that give rise to the observed dissociation trends. In Figure 1C, we show sampled configurations of the final DNA-bound structures of the various dimers. We immediately see that the thermodynamic binding differences are also accompanied by large-scale structural differences in the DNA-bound species.

By calculating the average elastic and electrostatic energy changes (see Materials and Methods) between the structural ensembles of the DNA-bound and the DNA-free NF κ B dimers, we see that the structural changes in the dimers upon DNA binding are accompanied by a changing balance between elastic energy and electrostatic energy. These two energies play off each other in a seesaw fashion. There is significant lowering in the internal elastic energy for the p65p50 heterodimer, which exceeds the energetic stabilization of the p65p65 and p50p50 homodimers (Figure 2). On the other hand, the electrostatic stabilization upon DNA binding appears comparable for p65p50 and p50p50 dimers but is less favorable for the p65p65 homodimer. Thus, both our calculations and the experiments point out that resolving the binding paradox requires one to look beyond the protein–DNA interaction surface itself and to inquire further into the way transcription factors are built to see how their inner elastic strains and the wiring of their contacts enables these proteins to sterically accommodate and release the DNA.

We therefore carried out long-time-scale (equivalent to $\sim 10^2$ μ s in laboratory time) constant-temperature simulations of the homo- and heterodimeric forms of NF κ B in their free and DNA-bound forms. We then analyzed the ensembles of conformations generated in this way using several types of PCA. PCA provides a way of decoupling correlated fluctuations to find dominant collective coordinates that may function as reaction coordinates in exploring conformational transitions.^{21,38,39} This analysis allows us to uncover how the accommodation of the protein to the DNA is made possible not only by rigid-body domain motions and localized strain rearrangements but also ultimately by cracking at specific sites in the dimer interfaces. These dimer interfaces apparently have evolved to be frustrated in the ground-state conformation so as to facilitate their rupture.

To disentangle these different collective motions, we employ several different forms of PCA. One of these analyses is the conventional PCA based on describing protein conformations using the Cartesian positions of residues in three-dimensional space.⁴⁰ Many papers^{8,21,38,39} have documented that these Cartesian principal components generally track the rigid-body motions of protein subdomains relative to each other.

Two other different kinds of PCA give further insight into the way these motions occur and allow us to identify the specific contacts whose local stretching and contraction contribute to the changes in elastic energy. We use one form employing pair distances in the space of physical contacts that shows where the molecule stretches and contracts (strain PCA) and a distinct form of analysis involving another set of observables that measures whether any contacts are completely made or broken (cracking PCA), which thereby monitors cracking (see Materials and Methods). Years ago, cracking PCA in contact space was used to predict and follow the subtle conformational changes induced by phosphorylation, a common form of post-translational modifications in proteins.^{33,41} Finally, by following both the dominant

conventional Cartesian principal components and the strain principal components, we find the specific locally strained contacts which are most correlated with the global structural changes in binding.

As visualized by Cartesian PCA, the dominant thermal motions of the free NF κ B dimers are domain twisting and breathing motions of the two DNA-binding domains relative to each other.⁸ These motions account for over 90% of the thermal fluctuations in the Cartesian coordinates of the protein (Supporting Information). The twist motion persists in the DNA-bound form, where it now appears as a shearing motion accompanied by one-dimensional sliding along the DNA. The breathing motion in the free protein dimers, on the other hand, is suppressed by binding DNA, which fills the cavity between the NF κ B binding domains. The suppression of the twist mode and the activation of the breathing mode, therefore, are seen as important contributors to the dissociation of NF κ B–DNA complexes. In a previous paper,⁸ we have shown how the naturally evolved inhibitor protein I κ B facilitates dissociation by freezing the domain twist in an open state. In this way, I κ B holds one of the domains of NF κ B in a twisted configuration so that one of the domains can no longer make strong electrostatic contact with the DNA, thereby allowing the DNA to escape the embrace of NF κ B more easily since it now makes contact with only a single NF κ B domain. In the absence of I κ B, the dissociation can only take place by means of energetically costly thermal fluctuations, which accomplish the domain twist and allow the breathing motion and thus DNA escape.

While the Cartesian principal components give a global picture of the motions needed for binding in each of the individual dimers, the energetic differences between these dimers come from changes that are localized in the protein structure (Figure 1C). These structural changes are captured by the local forms of PCA (Figure 3), which allow us to pin down those specific contacts that facilitate the global motions that are highlighted in the Cartesian principal components. The principal components of the fluctuations of the relative distances involved in local individual contacts localize the pivots for the large-amplitude motions of the molecule. We used these strain principal components to detect the regions most prone to deformation and cracking. Thus, when the strain becomes large enough, we see that some specific contacts break while others may form (Figures 4 and 5). The principal components of the correlated fluctuations of occupation of contacts thus detect and localize the sites of cracking; i.e., they locate the fault lines that initiate the “proteinquake”.

While globally the main principal Cartesian motions appear similar for all of the dimeric forms of NF κ B (Figure 3 and Supporting Information), the localized strain motions that drive these global motions are quite different for the heterodimer and for the homodimers. The strain principal components thus allow quantification of the type and magnitude of various contact contributions to the overall elastic energy changes (Figure 2). Additionally, using the correlations between Cartesian and strain principal components, we find that the twist results in a rather cooperative contact straining/breaking at the dimerization interface, while the breathing motion and lower amplitude thermal motions take place via a coordinated exchange of contacts, some contacts being made and others being broken more or less at the same time. This pattern is seen by computing the principal components of the binary contact probability fluctuations that monitor cracking (Supporting Information).

Projecting the conformational ensemble of the DNA-bound NF κ B complex onto the first strain principal component, we see that, upon binding to DNA, the multiple conformational states of the free p65p50 heterodimer NF κ B coalesce into one single macrostate (Figure 3). For the homodimers we see an opposite pattern, where the relatively more rigid dimer interface in the free dimer becomes more dynamic and the strain principal component distribution bifurcates into multiple states upon binding to DNA. This change in the strain PCs indicates that there is global frustration (i.e., frustration at a distance) between, on the one hand, the tendency of NF κ B dimers to form more stable rigid dimers through contacts in the dimerization interface and, on the other hand, their need to allow some conformational flexibility at the dimerization interface so as to be stronger DNA binders.

These local structural changes contribute toward the total elastic energy, which in the case of both homodimers is less favorable in the DNA-bound form than in the free form (Figure 2). The fluctuations of the first strain principal component in heterodimer p65p50 happen to be the most correlated with the global twist mode (Supporting Information). On the other hand, owing to the greater rigidity of the homodimers, the global twist mode is most correlated with a minor principal component, the 11th strain principal component, PC₁₁ (Supporting Information). This high-order principal component makes a factor of ~ 10 smaller contribution to the total fluctuations than does the first strain principal component. The first 10 strain principal components in both homodimers are uncorrelated with global motions of domains relative to each other but rather correspond to a variety of intradomain local deformations. This is why the total internal elastic energy of homodimers quantified by the quasi-harmonic approximation is much lower than it is for the p65p50 heterodimer (Figure 2). If one compares the average electrostatic free energy stabilization upon DNA binding, then electrostatically the p65p50 heterodimer appears to be comparable to p50p50. We see that what truly sets these dimers apart from each other is the nature of the elastic energy changes upon the DNA binding.

We therefore now turn to an analysis of the local frustration patterns in NF κ B dimers to understand why these local deformations are so different in these different dimeric molecules while their global motions may appear to be not too different from each other. In Figure 4, we show the frustration pattern of the native contact map along with the strain principal component and the corresponding contact occupation principal component that shows where there is cracking in the region of high strain (see also Supporting Information). We see that frustration, strain, and cracking are all quite well colocalized at a specific fault in the protein, much like the fault lines in seismology.

Using the PCA in the space of the pairwise distances between residues in physical contacts from constant-temperature equilibrium sampling allows us to easily identify the most strained contacts present in the free NF κ B dimers (Figure 4). These contacts form along the interface between the dimerization domains of NF κ B. These contacts, therefore, contribute significantly to the total changes in elastic energy during the accommodation of the DNA, as quantified by the quasi-harmonic approximations used by Miyashita et al.²⁴ (Figure 2). The dimer interface region also turns out to be more frustrated in the heterodimer than it is in either of the homodimers (Figure 4). The fault in the p65p50 heterodimer has apparently evolved to be easy to break. The more frustrated interfacial contacts of the heterodimer are

precisely localized (Figure 4) so as to favor the shear motions that enable the twist mode in both the free and the DNA-bound forms.

The role of frustration can be quantified by following the changes in the number of frustrated contacts along the twist-like motion for each dimer (Figure 5). We find that the heterodimer stands out as the only dimer that forms more contacts and becomes less frustrated at the end of the twist motion, while both the p65p65 and p50p50 homodimers break more contacts and become more frustrated as the dimer deviates from its equilibrium-averaged configuration so as to accommodate better the DNA. This shows that the high mobility of the heterodimer correlates with its unusually high binding affinity to the DNA. The accommodating motions are facilitated by the positioning of frustrated contacts near the dimerization interface region (Figure 4).

In both the free p65p65 and p50p50 homodimers, the protein dimerization interface is less frustrated (Figure 4 and Supporting Information) than in the heterodimer. Thus, the homodimers are indeed correspondingly more rigid, so that the twist motion is suppressed in these homodimers to a great extent compared to the heterodimer. The low frustration of the dimerization domain interfaces in homodimers likely follows from symmetry principles.^{42,43} Indeed, both directed evolution experiments and simulations have provided ample evidence that symmetric associations are energetically most stable and are therefore more likely to be discovered by evolution.⁴⁴ There are symmetry arguments for the apparent over-representation of homodimers in the functional protein universe.⁴⁵ Thus, we suspect that heterodimerization in NF- κ B has been selected by evolution to be the dominant form, thanks to its manifesting a more beneficial trade-off between thermodynamic stability and allosteric functional DNA binding that is needed to allow kinetic control of a large genetic broadcasting network.⁷

Finally, by computing the protein elastic energy for each conformational ensemble along the DNA dissociation coordinate and by monitoring the cracking contacts along with the strain contacts in the quasi-harmonic approximation, we see that the elastic energy is strongly correlated with the rupture of interdomain contacts in the NF- κ B as it pulls away from DNA (Figure 6). By comparing these events during passive DNA dissociation to what happens when the I κ B induces the dissociation of the DNA and strips the heterodimer p65p50 from DNA sites, we see that I κ B's catalytic action can be traced to reducing the large elastic barrier (Figure 6C), which allows DNA to dissociate with less breakage of the dimerization interdomain contacts.

CONCLUSION

In the present work, we have tried to illustrate the way the interactions between the dimerization interfaces of protein domains in transcription factors determine the nature of the allosteric motions when the complexes bind to DNA. The differences in the free energy landscapes of the binding and dissociation of heterodimers and homodimers from DNA reflect the conflicts between the elastic energies determined by the flexibility of dimerization interfaces of the transcription factors and the need for forming a strong electrostatic contact at the interface between the DNA and the DNA-binding domains of the protein complexes.

We have carried out long-time-scale simulations with a coarse-grained predictive protein +DNA force field showing the way DNA binding modulates the elastic deformations of the free transcription factors. Employing principal component analysis in the space of physical contacts along with the more conventional principal component analysis in the space of Cartesian coordinates of all backbone atoms allows us to understand how frustration at the dimeric interfaces controls motional flexibility by lubricating the fault that leads to the “proteinquake” necessary for DNA binding. The more frustrated interface of the heterodimer p50p65 leads to increased conformational disorder in the free form of the heterodimer, allowing p50p65 to bind more intimately to DNA than do either of the two homodimers, p65p65 and p50p50, which by having less frustrated and thus more rigid dimerization interfaces are prevented from optimally accommodating the DNA.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge financial support by the D.R. Bullard-Welch Chair at Rice University, Grant C-0016, and PPG Grant P01 GM071862 from the National Institute of General Medical Sciences.

References

1. Ptashne, M. A genetic switch: Gene control and phage.lambda. Blackwell Scientific Publications; Palo Alto, CA: 1986.
2. Ptashne, M., Gann, A. Genes & signals. Cold Spring Harbor Laboratory Press; New York: 2002.
3. Ghim C-M, Almaas E. BMC Syst. Biol. 2008; 2:94. [PubMed: 18980697]
4. Burger A, Walczak AM, Wolynes PG. Proc. Natl. Acad. Sci. U. S. A. 2010; 107:4016–4021. [PubMed: 20160109]
5. Bundschuh R, Hayot F, Jayaprakash C. J. Theor. Biol. 2003; 220:261–269. [PubMed: 12468297]
6. Potoyan DA, Wolynes PG. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:2391–2396. [PubMed: 24469814]
7. Potoyan DA, Wolynes PG. Phys. Rev. E. 2017; 96:052305. [PubMed: 29347696]
8. Potoyan DA, Zheng W, Komives EA, Wolynes PG. Proc. Natl. Acad. Sci. U. S. A. 2016; 113:110–115. [PubMed: 26699500]
9. Potoyan DA, Zheng W, Ferreira DU, Wolynes PG, Komives EA. J. Phys. Chem. B. 2016; 120:8532–8538. [PubMed: 27098223]
10. Phelps CB, Sengchanthalangsy LL, Malek S, Ghosh G. J. Biol. Chem. 2000; 275:24392–24399. [PubMed: 10825175]
11. Rackov G, Hernández-Jiménez E, Shokri R, Carmona-Rodríguez L, Mañes S, Álvarez-Mon M, López-Collazo E, Martínez-a C, Balomenos D. J. Clin. Invest. 2016; 126:3089–3103. [PubMed: 27427981]
12. Hoffmann A, Natoli G, Ghosh G. Oncogene. 2006; 25:6706. [PubMed: 17072323]
13. Siggers T, Chang AB, Teixeira A, Wong D, Williams KJ, Ahmed B, Ragoussis J, Udalova IA, Smale ST, Bulys ML. Nat. Immunol. 2012; 13:95–102.
14. Dembinski HE, Wismer K, Vargas JD, Suryawanshi GW, Kern N, Kroon G, Dyson HJ, Hoffmann A, Komives EA. Proc. Natl. Acad. Sci. U. S. A. 2017; 114:1916–1921. [PubMed: 28167786]
15. Fagerlund R, Behar M, Fortmann KT, Lin YE, Vargas JD, Hoffmann A. J. R. Soc. Interface. 2015; 12:20150262. [PubMed: 26311312]
16. Bergqvist S, Alverdi V, Mengel B, Hoffmann A, Ghosh G, Komives EA. Proc. Natl. Acad. Sci. U. S. A. 2009; 106:19328–19333. [PubMed: 19887633]

17. Onuchic JN, Wolynes PG. *Curr. Opin. Struct. Biol.* 2004; 14:70–75. [PubMed: 15102452]
18. Bryngelson JD, Wolynes PG. *Proc. Natl. Acad. Sci. U. S. A.* 1987; 84:7524–7528. [PubMed: 3478708]
19. Ferreiro DU, Hegler JA, Komives EA, Wolynes PG. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:3499–3503. [PubMed: 21273505]
20. Stein, DL. *Spin glasses and biology.* World Scientific; New York: 1992.
21. Zhuravlev PI, Papoian GA. *Q. Rev. Biophys.* 2010; 43:295–332. [PubMed: 20819242]
22. Parra RG, Schafer NP, Radusky LG, Tsai M-Y, Guzovsky AB, Wolynes PG, Ferreiro DU. *Nucleic Acids Res.* 2016; 44:W356–W360. [PubMed: 27131359]
23. Kitao A, Takemura K. *Curr. Opin. Struct. Biol.* 2017; 42:50–58. [PubMed: 27815984]
24. Miyashita O, Onuchic JN, Wolynes PG. *Proc. Natl. Acad. Sci. U. S. A.* 2003; 100:12570–12575. [PubMed: 14566052]
25. Ansari A, Berendzen J, Bowne SF, Frauenfelder H, Iben I, Sauke TB, Shyamsunder E, Young RD. *Proc. Natl. Acad. Sci. U. S. A.* 1985; 82:5000–5004. [PubMed: 3860839]
26. Whitford PC, Miyashita O, Levy Y, Onuchic JN. *J. Mol. Biol.* 2007; 366:1661–1671. [PubMed: 17217965]
27. Whitford PC. *Proc. Natl. Acad. Sci. U. S. A.* 2013; 110:7114–7115. [PubMed: 23610426]
28. Shan Y, Arkhipov A, Kim ET, Pan AC, Shaw DE. *Proc. Natl. Acad. Sci. U. S. A.* 2013; 110:7270–7275. [PubMed: 23576739]
29. Li W, Wolynes PG, Takada S. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:3504–3509. [PubMed: 21307307]
30. Okazaki, K-i, Koga, N., Takada, S., Onuchic, JN., Wolynes, PG. *Proc. Natl. Acad. Sci. U. S. A.* 2006; 103:11844–11849. [PubMed: 16877541]
31. Okazaki, K-i, Takada, S. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:11182–11187. [PubMed: 18678900]
32. Ikeguchi M, Ueno J, Sato M, Kidera A. *Phys. Rev. Lett.* 2005; 94:078102. [PubMed: 15783858]
33. Lätzer J, Shen T, Wolynes PG. *Biochemistry.* 2008; 47:2110–2122. [PubMed: 18198897]
34. Davtyan A, Schafer NP, Zheng W, Clementi C, Wolynes PG, Papoian GA. *J. Phys. Chem. B.* 2012; 116:8494–8503. [PubMed: 22545654]
35. Freeman GS, Hinckley DM, Lequieu JP, Whitmer JK, de Pablo JJ. *J. Chem. Phys.* 2014; 141:165103. [PubMed: 25362344]
36. Lequieu J, Córdoba A, Schwartz DC, de Pablo JJ. *ACS Cent. Sci.* 2016; 2:660–666. [PubMed: 27725965]
37. Martone R, Euskirchen G, Bertone P, Hartman S, Royce TE, Luscombe NM, Rinn JL, Nelson FK, Miller P, Gerstein M, Weissman S, Snyder M. *Proc. Natl. Acad. Sci. U. S. A.* 2003; 100:12247–12252. [PubMed: 14527995]
38. Ernst M, Sittel F, Stock G. *J. Chem. Phys.* 2015; 143:244114. [PubMed: 26723658]
39. Ernst M, Wolf S, Stock G. *J. Chem. Theory Comput.* 2017; 13:5076–5088. [PubMed: 28915045]
40. Amadei A, Linssen A, Berendsen HJ. *Proteins: Struct. Funct. Genet.* 1993; 17:412–425. [PubMed: 8108382]
41. Shen T, Zong C, Hamelberg D, McCammon JA, Wolynes PG. *FASEB J.* 2005; 19:1389–1395. [PubMed: 16126906]
42. Wolynes PG. *Proc. Natl. Acad. Sci. U. S. A.* 1996; 93:14249. [PubMed: 8962034]
43. Blundell TL, Srinivasan N. *Proc. Natl. Acad. Sci. U. S. A.* 1996; 93:14243–14248. [PubMed: 8962033]
44. Plaxco KW, Gross M. *Curr. Biol.* 2009; 19:R25–R26. [PubMed: 19138586]
45. Bergendahl LT, Marsh JA. *Sci. Rep.* 2017; 7:4932. [PubMed: 28694495]

Figure 1.

(A) Simplified kinetic scheme that depicts the kinetics of interconversion between DNA-bound, free, and IκB-bound NFκB dimers. Due to a combination of factors including large number of binding sites and, on average, higher affinities, the p65p50 heterodimer of NFκB relies on molecular stripping for generating a timely response to appropriate external signals. For the p65p65 and p50p50 homodimers, either there is no molecular stripping (p50p50 lacks NLS domain) or any effect of stripping is negligible because of the low binding affinity (p65p65). The nearly irreversible stripping process indicated by the red arrow renders the heterodimer population under kinetic control. (B) Free energy profiles of DNA dissociation from different NFκB–DNA complexes. The p50p65 heterodimer is shown in blue, p65p65 homodimer is shown in green, and p50p50 homodimer is shown in red. (C) Representative structures from the simulations of NFκB dimers bound to the same 20-bp-long DNA segment.

Figure 2.

(A) Differences in average protein elastic energies between DNA-bound and free forms of NF κ B: $E_{\text{elast}}(\text{prot:DNA}) = E_{\text{elast}}(\text{prot:DNA}) - E_{\text{elast}}(\text{prot})$. (B) Differences in average electrostatic energies (estimated at a mean-field Debye-Hückel level) between the DNA-bound and free forms of NF κ B: $E_{\text{elec}}(\text{prot:DNA}) = E_{\text{elec}}(\text{prot:DNA}) - E_{\text{elec}}(\text{prot})$.

Figure 3.

Distributions of Cartesian and strain PCs for all three dimers of NF κ B in their free and DNA-bound structural ensembles. (A) Distribution of twist and first strain PC for p65p65 and p65p65–DNA. (B) Distribution of twist and first strain PC for p65p50 and p65p50–DNA. (C) Distribution of twist and first strain PC for p50p50 and p50p50–DNA.

Figure 4.

Contact maps color-coded to show the frustration patterns (lower triangle) and the main contacts that contribute to strain motion (upper triangle) (PC_1) for p65p650, p65p50, and p50p50 dimers. The color bar at the right indicates the magnitude of fluctuation in strain PC motion. The color of the contacts for the frustration patterns shown in both the contact map and the three-dimensional structures is indicated as red for highly frustrated contacts and green for minimally frustrated contacts.

Figure 5.
Change in the frustration along the twist modes. (A) Change in the total number of all contacts. (B) Change in the total number of frustrated contacts as the dimeric molecules traverse along the twist coordinate.

Figure 6.

Profiles of the elastic energy along the DNA dissociation coordinate for DNA dissociation from $I\kappa B$ -bound (red) and $I\kappa B$ -free (blue) heterodimer $p50p65$ forms. (A) Change in the total elastic energy of $NF\kappa B$. (B) Elastic energy associated with the cracking PC_1 , which is localized at the dimerization interface of the $p65p50$ heterodimer. (C) Structures that correspond to transition states in the cracking PC_1 profile, which monitors making/breaking of contacts at the dimerization interface. Contacts between the dimerization domains are highlighted, with red lines indicating cracked contacts ($q_{ij} = 0$) and blue ones indicating uncracked ones ($q_{ij} = 1$).