Title
Allosteric Autoinhibition Pathway in Transcription Factor ERG: Dynamics Network and Mutant Experimental Evaluations

Journal
Journal of Chemical Information and Modeling, 57(5)

ISSN
1549-9596

Authors
Ye, Wei
Qian, Tianle
Liu, Hao
et al.

Publication Date
2017-05-22

DOI
10.1021/acs.jcim.7b00073

Peer reviewed

HHS Public Access

Author manuscript

J Chem Inf Model. Author manuscript; available in PMC 2017 December 14.

Published in final edited form as:

J Chem Inf Model. 2017 May 22; 57(5): 1153–1165. doi:10.1021/acs.jcim.7b00073.

Allosteric Autoinhibition Pathway in Transcription Factor ERG: Dynamics Network and Mutant Experimental Evaluations

Wei Ye[1,§], **Tianle Qian**[1,§], **Hao Liu**[1,§], **Ray Luo**[2,*], and **Hai-Feng Chen**[1,3,*]

[1]State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, SJTU-Yale Joint Center for Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, 200240, China

[2]Departments of Molecular Biology and Biochemistry, Chemical Engineering and Materials Science, Biomedical Engineering, University of California, Irvine, California 92697-3900, USA

[3]Shanghai Center for Bioinformation Technology, 1278 Keyuan Road, Shanghai, 200235, China

Abstract

Allosteric autoinhibition exists in many transcription factors. The ERG proteins exhibit autoinhibition on DNA binding by the C-terminal and N-terminal inhibitory domains (CID and NID). However, the autoinhibition mechanism and allosteric pathway of ERG are unknown. In this study we intend to elucidate the residue-level allosteric mechanism and pathway *via* a combined approach of computational and experimental analyses. Specifically computational residue-level fluctuation correlation data was analyzed to reveal detailed dynamics signatures in the allosteric autoinhibition process. A hypothesis of "NID/CID binding induced allostery" is proposed to link similar structures and different protein functions, which is subsequently validated by perturbation and mutation analyses in both computation and experiment. Two possible allosteric autoinhibition pathways of L286-L382-A379-G377-I360-Y355-R353 and L286-L382-A379-G377-I360-Y355-A351-K347-R350 were identified computationally and were confirmed by the computational and experimental mutations. Specifically we identified two mutation sites on the allosteric inhibition pathways, L286P/Q383P (NID/CID binding site) and I360G (pathway junction), which completely restore the wild type DNA binding affinity. These results suggest that the putative protein structure-function relationship may be augmented with a general relationship of protein "structure/fluctuation-correlation/function" for more thorough analyses of protein functions.

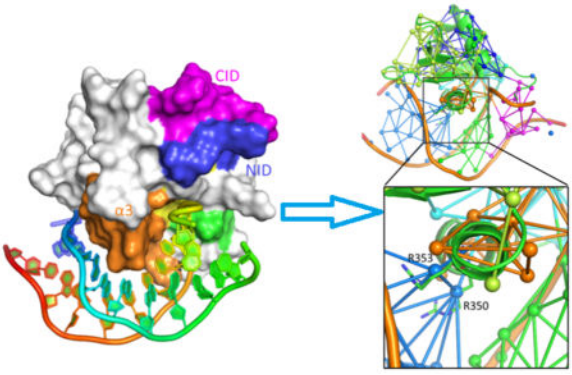
Graphical Abstract

*Corresponding authors: haifengchen@sjtu.edu.cn; rluo@uci.edu, Tel: 86-21-34204348, Fax: 86-21-34204348.
§These authors contributed equally to this work.

The authors declare that there is no conflict of interest.

Supporting Information

Figure S1 plots the convergence trend lines of the $\Delta\Delta G$ for ERGu (Y354F) and ERGi (Y354F). Figure S2 plots purification and plasmon resonance analysis for ERG protin. (a) Coomassie-stained 15% SDS-PAGE of purified ERG protein. Lines: 1. Maker; 2. His-ERGu (12.9 KD); 3. His-ERGi (15.8 KD) B-F. Surface plasmon resonance analysis of corresponding plot of steady state response against concentration for determination of binding affinity. (b) ERGu; (c) ERGi; (d) ERGu R353A; (e) ERGu Y354F; (f) ERGi R353A; (g) ERGi Y354F; (h) ERGi I360G; (i) ERGi L286P/Q383P. The vertical line indicates the value of the calculated equilibrium dissociation constant K_D. Figure S3 plots linear correlation between K_D derived from SPR measurements and ΔG from MM/GBSA. This material is available free of charge via the Internet at http://pubs.acs.org.



Introduction

Autoregulation refers to the protein function regulated, inhibited, or enhanced by itself via internal responses to binding of its functional substrate or triggering of its co-regulatory partner [1–3]. It is an economic way for proteins to regulate their activities or their substrates and is observed in many important proteins, such as transcription factors [1, 4, 5], receptors [6], and kinases [7, 8]. In many autoregulation processes allostery is found, i.e. the regulatory site is at a physically distinct site away from its active site of the same protein [5, 9]. Recently the roles of autoregulation and its allosteric mechanisms in biochemical processes have drawn widespread attention [8, 10, 11]. Proteins with autoinhibition, i.e. negative autoregulation, are also promising drug targets. Several drugs designed by targeting autoinhibited proteins have been reported [12–14].

In typical allosteric regulations with small molecules as allosteric regulators, structural changes are often observed [15, 16]. An allosteric regulator binds to a protein of interest and induces a structural change, which leads to a functional change of the protein. Based on this principle [17], pathways are identified in several allosterically regulated proteins [18]. However, little structural changes could be observed in autoregulation processes [10, 19] so that many autoregulation mechanisms and their relationships to structure and function are hard to decipher.

In this study, we focus on the allostery mechanism in the autoinhibition of an ETS transcription factor family protein, ERG, upon binding to its target DNA. The ERG proteins are unique to animals [20]. They recognize a central GGA motif on DNA and play key roles in the regulation of megakaryocyte differentiation [2] and vasculogenesis/angiogenesis [21]. Indeed, the *ERG* gene's fusion with the *TMPRSS2* gene is found in over 50% of the prostate cancer [22, 23]. The fusion leads to overexpression of *ERG* in the prostate epithelium. Furthermore, fusions of *ERG* were also found in acute amyloid leukemia [24] and Ewing's sarcoma [25], indicating the importance of *ERG*'s regulations. Recently X-ray crystallography and EPR spectroscopy studies indicate the presence of allosteric autoinhibition of ERG upon binding with its target DNA. More interestingly, there is little conformational change in its core ETS domain [19]. The autoinhibition effect only exists when both N-terminal inhibitory domain (NID) and C-terminal inhibitory domain (CID) are present; it disappears when NID and CID are absent. Allosteric autoinhibition is also observed in other ETS proteins [26–28].

Crystal structures of holo- and apo-ERG proteins provide valuable insights into the autoinhibition allostery mechanism [19]. As shown in the crystal structures (Fig. 1a), conserved core domain binds with the DNA GGA motif at the interface of ERG's helix α3 and the DNA molecule's major groove. NID and CID, located at the two terminal regions of the core domain, are in close contact with each other. A long hairpin loop (L1) also has a large interaction interface with the DNA. The sequence and secondary structure of these domains are also illustrated in Fig. 1b with intrinsically disordered regions. An interesting observation is that conformational difference between bound autoinhibited ERG (holo-ERGi) and bound uninhibited ERG (holo-ERGu) is well within the thermal noise (all-atom RMS*d* of 0.35 Å) (Fig. 1c). The Bushweller group subsequently proposed that the ERG autoinhibition regulation is, among several other hypotheses, predominantly mediated by the protein dynamics due to the absence of noticeable structural changes [19].

Thus a natural question is whether dynamic features can be used to explain the distinct functional differences between autoinhibited and uninhibited states of the ERG protein. To answer these questions, dynamical networks, which were previously used to illustrate the allosteric behavior of aminoacyl-tRNA synthetase [17] and several other systems [29, 30], were constructed based on all-atom molecular dynamics (MD) simulations of the wild type and mutant ERG protein. From the comparisons of the networks between holo-ERGi and holo-ERGu, a hypothesis of "NID/CID binding induced allostery" and two possible allosteric autoinhibition pathways were proposed to link similar structures and different protein functions, which are subsequently validated by both computational and experimental mutation analyses.

Results and Discussion

Dynamical Features in Different ERG-DNA Complexes

Both crystal structures and modeled structures show relatively high stabilities after equilibration in MD simulations. Root mean square deviation (RMS*d*) with respect to the initial structure (Fig. 2) shows that 150 ns simulations are sufficient for equilibration in most wild type and mutant systems. For the deletion mutants, holo-ERGi(NID-) and holo-ERGi(CID-), 200 ns were simulated due to the large structural perturbation to the crystal structure. The average structures (over the last 50ns of all trajectories) of the core DNA binding domains in holo-ERGi and holo-ERGu were aligned and shown in the insert of Fig. 3a. The average all-atom RMS*d* is only 0.72 Å, showing little structural differences between autoinhibited and uninhibited core DNA-binding domains.

The root mean square fluctuations (RMS*f*) of Cα (in protein) or C5' (in DNA) are also shown in Fig. 3a for apo-ERGu, apo-ERGi, holo-ERGu, and holo-ERGi, respectively. RMS*f*'s for single-point mutated and truncated complexes are shown in Fig. 3b. In general, the Cα variations in the core domain of holo-ERG are slightly lower than those of the apo-ERG. As expected in typical MD simulations, large fluctuations are focused on the N and C terminal regions for these systems.

In order to confirm the reliability and robustness of the MD simulations, chemical shift perturbations between holo-ERGu and holo-ERGi are calculated and shown in Fig. 4. This

figure indicates that the results of MD agree well with Bushweller group's ^{15}N-^{1}H heteronuclear single-quantum coherence (HSQC) spectra. Both experiment and computation suggest that significant change regions of chemical shift are also located on the N and C terminal domains. However these local conformational changes do not reveal how they are responsible for the observed differential DNA binding.

Dynamical Networks of Autoinhibited and Uninhibited ERG Proteins

Both experiment and computation indicates that there are significant chemical shift differences in the N- and C-terminal domains of the protein. However, these changes do not provide a molecular basis for the difference in DNA binding [19]. In order to reveal the molecular mechanisms responsible for the different binding affinities, we analyzed the residue-level fluctuation correlation matrix to identify essential dynamics signatures that are in play in autoinhibition and allostery.

A revised dynamical network analysis method was utilized to study the fluctuation correlation information. To understand the motivation of the new strategy, envision each residue in a protein as a person in a society, so that its fluctuation is mutually influenced by other contacting residues just like the person's conversation is mutually influenced by other persons in contact. If we treat each residue/person as a node in a network, all nearby nodes are correlated as far as information flow is concerned. Therefore, information flow analytic properties are useful in understand a network whether it's a social network or a fluctuation correlation network.

To construct a dynamical network on fluctuation correlation, the covariance matrices (C_{ERGi} and C_{ERGu}) on fluctuation correlation were first calculated for holo-ERGi and holo-ERGu and are shown in Fig. 5, which shows that the main differences between holo-ERGi and holo-ERGu are in the binding interface. Based on the covariance matrices for the complexes, correlation networks of holo-ERGi and holo-ERGu were constructed [17]. In order to reduce the effect of thermal fluctuations in any single trajectory, each network was constructed using the conformations sampled from three independent trajectories.

The basic network topology characterizations, analyzed as in Materials and Methods, are gathered in Table 1. The network centralization, density, and heterogeneity of holo-ERGu are higher than those of holo-ERGi. The average numbers of neighbors for the holo-ERGu network are also higher than those of the holo-ERGi network. These indicate that the dynamical networks of holo-ERGu are apparently different from those of holo-ERGi.

Figure 6 shows the dynamical networks for both holo-ERGu and holo-ERGi. In addition it also highlights network hub nodes (labeled by the residues/subunits), i.e. the proportion of correlation weighted degrees (CWDs) for the node is larger than 1% of the CWDs for the whole nodes. Worth noting, there are 38 hub nodes in the holo-ERGu network and 33 such nodes in the holo-ERGi network, for example Y354 in both systems and R350 and R353 in holo-ERGu. These hub nodes play key roles in the information transfer of network, so the ease of information transfer is higher in the holo-ERGu network. Additionally more hub nodes are concentrated on the DNA binding interface in the holo-ERGu network. This

suggests that information transfer is particularly high on the DNA binding interface in the holo-ERGu network.

Figure 7 further shows that more nodes are with higher CWDs on the DNA binding interface in the holo-ERGu network, supporting the conclusion that information transfer is higher in the holo-ERGu network, particularly on the binding interface.

Network Differences Correlate with Binding Affinity Differences

Given that the dynamical networks of autoinhibited holo-ERGi and uninhibited holo-ERGu are different, it is natural to ask whether the network differences can be used to interpret their different binding affinities. For a specific information network, variations in the connectivity of the network give rise to local communities. A node is more strongly and densely connected to those in the same community than to those in a different community. Community analysis may better account for the variations in the information flow among structural components in a network [17]. We utilized the Girvan-Newman algorithm to split the complex networks into multiple communities, and then analyzed the differential community connectivity between autoinhibited and uninhibited complex networks.

The community networks are shown in Fig. 8 with nodes and communities projected onto the 3D structures. The analysis shows that distinct differences between holo-ERGi and holo-ERGu were mainly found at two binding interfaces, which is consistent with the networks' differences. *1) The N-terminal region* In holo-ERGu (uninhibited), the target DNA is partially classified into the same community with the protein's N-terminus. While in holo-ERGi (autoinhibited), the N-terminus and the target DNA are classified into two different communities with no connecting edges. This indicates that the interaction between the N-terminus and the DNA is weaker in holo-ERGi than in holo-ERGu. *2) Helix α3 and the DNA's major groove* In the holo-ERGi network, no edges between helix α3 and DNA are found. This indicates that the information flows between these two components are very weak. Consequently, the information from one component to the other must transfer through other components without direct pathway. In contrast several edges exist at the second binding interface in the holo-ERGu network. Two arginines (R350 and R353) on the helix α3 are in contact with the major groove of DNA (Fig. 8), which indicates that these two residues are very important for the binding interface. Interestingly, R353 is adjacent to Y354, a key functional residue identified in a previous experiment [19]. Even though both are on the binding interface, our network community analysis shows they have different network properties, which correlate with their different binding. However, this difference cannot be understood just by analyzing the structures of holo-ERGu and holo-ERGi alone.

NID/CID Binding Induces Allostery Responsible for Differential Information Flows and Binding Affinities

Our analysis shows that dynamical networks are significantly different between holo-ERGi and holo-ERGu. Furthermore, the differential information flows at the binding interface correlate well with the relative binding affinities. Since the structures of the binding domain in the two complexes are highly similar, the network differences must be caused by the presence of NID and CID. However, the binding interface, helix α3 and the DNA major

groove are not close to NID or CID (Fig. 9a). Thus autoinhibition due to the presence of NID or CID is most likely to be allosterically regulated, as pointed out by the Bushweller group[19]. They further proposed that the allostery is not driven by conformational alterations as the changes observed in the crystal structure were quite subtle particularly at the binding interface, as we also observe in our MD simulations.

Further structural analysis suggests that S284, L286 in NID, and Q383 in CID form a hydrogen-bonding network between NID and CID, as shown in Fig. 9b. Therefore, a "binding-induced allostery" mechanism may be proposed to explain the effect of NID or CID upon the allosteric autoinhibition. In order to validate this mechanism, we tried to weaken or destroy the interactions between NID and CID to study their influences upon the dynamical network and binding affinities as follows.

In previous experiments, CID was reported to be more important than NID as an inhibitory cassette: In ETV6, a transcriptional repressor in the ETS family, only CID is found to inhibit the DNA binding [32, 33]. Thus we first removed CID to build a deletion mutant complex, holo-ERGi(CID-). After 200 ns simulation, dynamical network was constructed as other complexes. In the network community analysis, helix α3 and the DNA major groove are merged into the same community (Fig. 10a), indicating that the two structural components become strongly correlated. Inspection of the networks of holo-ERGi(CID-) and holo-ERGu (Fig. 8b) show that the two networks are more similar: both with more connections thus more information flow on the binding interface. This is correlated with a more favorable binding interaction, consistent with previous observations, as shown in Table 2 [19]. In summary, both experimental and computational studies show that the removal of CID, upon destroying the interactions between NID and CID leads to stronger binding, even if the binding site is not close to CID. This structural perturbation and its consequences confirm the hypothesis that binding between NID and CID potentially induces an allosteric regulation.

Second, NID was removed from holo-ERGi to build deletion mutant holo-ERGi(NID-) to support the hypothesis, similar to the analyses for holo-ERGi(CID-). As Fig. 10b shows, the mutant's community splitting is similar to that of holo-ERGu, with R353 on helix α3 in contact with DNA, distinctly different from holo-ERGi (Fig. 8a) without any connections on the binding interface. In holo-ERGi(NID-), information can transfer from protein to DNA (or vice versus) directly through R353 on the binding interface, instead of through other components in holo-ERGi, leading to higher level of information flow at the binding interface, which correlates with a higher binding affinity (Table 3). Experimentally this system also shows a decrease of inhibition with K_D of 41 nM [19], supporting our computational analysis (Table 2).

Third, we weakened the binding between NID and CID by setting the correlation-weighted degrees between two domains as zero while all the other components in the dynamical network remain unchanged. The network community of holo-ERGi (Fig. 10c) upon the hypothetical network alteration shows similar results as holo-ERGu. Helix α3 is now connected with the DNA major groove, which indicates that information flows between the two become easier at the binding interface when NID and CID do not interact with each

other, thus leading to higher binding affinity. These observations confirm the interactions between NID/CID may perturb the DNA binding.

Finally, according to the interaction analysis, we attempted to destroy the interactions between NID and CID by conducting a double-mutation of L286P/Q383P on holo-ERGi, as L286 and Q383 form key backbone hydrogen bonds and native contacts. The network and community of this double-mutant system (Fig. 10d) are similar to those of holo-ERGu, where R350 and R353 are part of the communities nearby the DNA major groove, indicating that the two arginines on helix α3 are closely connected with the DNA. This is significantly different from the network of holo-ERGi, where none of the protein components and DNA are clustered into the same community. Indeed our subsequent thermodynamic integration simulations show that the binding affinity of the double-mutant is more favorable (−2.87 kcal/mol, see Table 2) than that of WT (holo-ERGi). The higher binding affinity was also confirmed in our surface plasmon resonance (SPR) experiment, with an 8.59 fold lower K_D than holo-ERGi (Table 2). More importantly, the affinity of the double mutant is already at the level of holo-ERGu, showing almost complete restoration of function in the holo-ERGi system. Thus both computational and experimental mutational analyses confirm that the destruction of the binding interface between NID and CID may change the network and information flow, and change the binding affinity favorably. At the same time, we decomposed the binding free energy for R285 that the value R285 is about −4.03 ± 2.75 kcal/mol in ERGi, and −6.32 ± 2.69kcal/mol in ERGi(L286P/Q383P). This suggests that R285 also plays a key role in DNA binding.

In summary, uninhibited ERGu-like dynamical network and information flow emerge on autoinhibited holo-ERGi complex when it undergoes deletion mutation (CID or NID removed), the double-point mutation (L286P/Q383P), and the hypothetical network alteration. These network changes lead to reduced inhibition upon DNA binding. Because the changes on network and edge strength are due to the destruction of the NID/CID binding, it can be concluded that the NID/CID binding interaction plays a key role in changing the network information flow, and inducing the allostery. Subsequently, the allostery causes the changes in DNA binding affinity.

NID/CID Binding-Induced Allostery Pathways

The Bushweller's group previously proposed a correlation between differential protein dynamics and allosteric inhibition, among several other hypotheses, given the absence of large conformational changes [19]. Our detailed all-atom molecular dynamics simulation and dynamical network analysis lead the hypothesis of binding-induced allostery. Our subsequent analyses with weakened, deleted, and mutated NID and CID support the hypothesis.

The next issue is the actual allostery pathway. Shortest path algorithm [34] was used to identify the pathway from the NID-CID binding (regulatory) site to the DNA binding (allosteric) site. Dynamical network and community analysis indicate that R350 and R353 are the most significant difference: the network community changes between holo-ERGu and holo-ERGi always occur on these two arginines (Fig. 8 and Fig. 10). Thus they may be the key nodes of information transfer at the allosteric site and are used as the destination of

the allostery pathways. Indeed, two allostery pathways in holo-ERGi are identified, L286-L382-A379-G377-I360-Y355-R353 and L286-L382-A379-G377-I360-Y355-A351-K347-R350 with R350 and R353 as end nodes, respectively (Fig. 11). In summary, we can describe the pathways as: NID/CID binding interface → Loop linking α3 and β3 → α3. Moreover, we failed to find a similar pathway in holo-ERGu, suggesting no inhibition exists in holo-ERGu.

Hypothetical network alteration was used to validate the two allosteric pathways. Note that Y355 is located on the junction of the two pathways and is thus crucial for the information flow, but it is also important for the structural integrity of the binding interface and its mutation unfolds the local structure. Instead, we mutated I360 to confirm the validity of the pathways. Without carrying out another MD simulation, the alteration was realized by setting the strength of all I360's edges as zero in the network. The community analysis shows a significantly repartitioned network (Fig. 11b) that is similar to that of holo-ERGu: R350 is clustered into one community with the DNA major groove, so that the information flow may be transferred directly at the binding interface, instead of being detoured in holo-ERGi (Fig. 8a). Thus the alteration leads to higher information flow and higher affinity, which could be further confirmed by $\Delta\Delta G$ of −0.63 kcal/mol (TI calculation) and 4.48 fold lower K_D (SPR experiment) upon the I360G mutation (Table 2). The linear correlation of $\Delta\Delta G$s for different systems between experiment and computation is shown in Fig. 12. A high correlation coefficients of 0.96 from TI (Figure S3) and 0.88 from MM/GBSA (Figure S4) show a strong correlation between the simulation and the experiment. Again, its DNA binding affinity is restored to the level of holo-ERGu, showing almost complete restoration of function in the holo-ERGi system. This suggests that the proposed allostery pathways, at least I360 before the junction, play key roles in allosteric autoinhibition.

Comparison with Previous NMR experiments

The structural analysis has shown that Lys347, Arg350, Arg353, and Tyr354 of ERGu are key residues to DNA interaction[19]. Eight stable electrostatic interactions were found for Arg350/DNA, Lys347/DNA, Arg350/DNA, and Arg353/DNA in our room temperature simulation. Furthermore, two stable hydrogen bonds was formed for Lys347/DNA and Tyr350/DNA. These results are in good agreement with the structure analysis that Lys347, Arg350, and Arg353 form important interactions with DNA. The previous work also reports that Y354F mutant greatly weakened the binding of ERGu with DNA. The electrostatic interactions were decreased to 26 from 31 (ERGu wild type) and just one hydrogen bond remaining. These results are also consistent with the NMR experiment.

Alternative Mechanisms to the ERG-DNA Binding Regulation

In this study residue-level dynamical network analyses of residue-level fluctuation correlation data assisted the formulation of the "binding-induced allostery" hypothesis and helped the identification of the allosteric autoinhibition pathway in ERG DNA-binding. The computational free energy simulation and experimental SPR measurements support the hypothesis and allosteric pathway.

The previous work suggests that the molecular mechanism of ERG autoinhibition is the combination of subtle conformational adjustment and changes in protein dynamics[19]. Specifically, two rotameric states for Y354 were identified in the crystal structures: in the first state it is positioned to hydrogen bond with specific bases in the target DNA; in the second state it hydrogen bonds to Ser283 in the NID[19]. In addition the chemical shift of Y354 is significantly changed between ERGu vs ERGi[19]. Finally upon the S283A mutation, the DNA binding affinity is also partially restored from that of the ERGi[19].

Our molecular dynamics network analysis shows that the end point of the allosteric pathway is R350 and R353, with the latter covalently bonded to Y354. All these residues are located on the α3 DNA binding interface. They all strongly correlate with the target DNA molecules as shown in Table 3, though with R350 and R353 slightly higher. Thus our network analysis also supports the important role of Y354, which is part of the allosteric pathway (which includes all directly covalent-bonded neighbors to the on-pathway residues shown in Fig. 11). In addition, mutation Y354F reduces DNA binding[19] by 3.1 fold in ERGu and by 1.4 fold in ERGi (Table 2), similar to the effect of mutation Y353A, which leads to 3.4 fold reduction in ERGu and 2.2 fold reduction in ERGi (Table 2). These values are consistent with those reported in the literature[19].

Our detailed molecular dynamics/network analysis facilitates the formulation of a more detailed hypothesis, leading to the identification of the residue-level pathway starting from the CID/NID binding interface. Our computational and experimental mutational analyses confirm that the destruction of the binding interface between NID and CID may change the network and information flow, and change the binding affinity favorably. Specifically, both mutation L286P/Q383P on the CID/NID binding interface and on-pathway mutation I360G were found to restore ERGi DNA binding affinities to the original wild type ERGu level (Table 3).

The flexibility of the NID was also proposed to be important for the autoinhibition of ERG protein[19]. Its flexibility allows it to be displaced entirely as a result of DNA binding as shown in the crystal structure[19]. In our molecular dynamics network analysis, removal of NID from ERGi was found to lead to a system more connected between the protein and the target DNA. These results support the hypothesis of NID/CID-binding-induced allosteric autoinhibition and are in agreement with the previous works[19]. It was also hypothesized that NID might interact with other binding partners as in the cooperative binding of the transcription factors ETS-1 and Pax5 to DNA [6–12, 15–19]. Our analysis complements this possible mechanism on how such interactions may work, i.e. the additional intramolecular interactions might reduce the NID/CID interaction to achieve their regulation roles.

Materials and Methods

Computational structure preparation

Initial atomic coordinates of all complexes were retrieved from the Protein Data Bank: uninhibited apo-ERG without CID and NID (apo-ERGu, 289-378) was extracted from 4IRG; autoinhibited apo-ERG with CID and part of NID (apo-ERGi, 283-384) was extracted from 4IRH; and the autoinhibited holo-ERG (holo-ERGi, 283-385) was extracted from 4IRI.

Because the NID structure is incomplete in 4IRI, the nine missing residues (283-291) were modeled with 4IRH as template. Based on the structural similarity between uninhibited and autoinhibited apo-ERG (RMS*d*: 0.35 Å), the initial uninhibited holo-ERG structure (holo-ERGu, 289-378) was modeled with the autoinhibited holo-ERG from 4IRI as template. Besides these wild-type complexes, mutants R353A, Y354F, L286P/Q383P, and I360G were constructed. Deletions by removing either NID or CID were also modelled. All the modelling, mutations, and representations were conducted in PyMOL 1.7 [35]

Molecular dynamics simulations

All initial structures were first minimized in SYBYL®-X 2.1.1 [36] to eliminate any possible overlaps or clashes. All simulations and most analysis procedures were conducted using the AMBER12 software package [37]. Hydrogen atoms were added using the LEaP module of AMBER12. Counter-ions were used to maintain system neutrality. All systems were solvated in a truncated octahedron box of TIP3P waters with a buffer of 10 Å. Particle Mesh Ewald (PME) [38] was employed to treat long-range electrostatic interactions with the default setting in AMBER12. The *ff99IDPs* force field [39–41] was used for all simulations. The SHAKE algorithm [42] was used to constrain bonds involving hydrogen atoms. All the MD simulations were accelerated with the CUDA version of PMEMD [43, 44] and run on NVIDIA® Tesla K20. Up to 20000-step steepest descent minimization was performed to relieve any further structural clash in the solvated systems. This was followed by a 400-ps heating up and a 200-ps equilibration in the NVT ensemble at 298K. The heating and equilibration runs were simulated with a time step of 2 fs in the Langevin thermostat with default settings. Finally the production runs were simulated in the NPT ensemble at 298K with a time step of 2 fs in the Berendsen's thermostat and barostat with default settings.

Three independent production trajectories of 150 ns each were simulated for apo-ERGi, apo-ERGu, holo-ERGi, and holo-ERGu, respectively. 150 ns simulations were found sufficient for these systems to reach equilibrium at the room temperature. For the deletion mutants, holo-ERGi(NID-) and holo-ERGi(CID-), 200ns were simulated due to the large structural perturbation to the experimental structure. 2.95 μs trajectories in total were collected for all simulated systems including the wild type and mutant proteins, taking about 1,100 GPU hours. Detailed simulation conditions are listed in Table 4. All structures, both from experiment and modelling, show relatively high stabilities after equilibration in MD simulations. RMS*d* with respect to the initial structure (see Fig. 2) shows that 150 ns simulations are sufficient for most wild type and mutant systems.

Dynamical Network Analysis

Dynamical fluctuation correlation network is constructed from the atomic fluctuation covariance matrix [45–48]:

$$C_{mn}=\frac{\langle \Delta\vec{r_m}(t)\cdot\Delta\vec{r_n}(t)\rangle}{\sqrt{(\langle \Delta\vec{r_m}(t)^2\rangle\langle \Delta\vec{r_n}(t)^2\rangle)}} \quad (1)$$

where displacement of node m is computed respect to its time average as $\Delta\overrightarrow{r_m}(t)=\overrightarrow{r_m}(t)-\langle\overrightarrow{r_m}(t)\rangle$, $\langle\cdot\rangle$ is the time average of node m, and $\overrightarrow{r_m}(t)$ is the position of node m. Fluctuation sampling data was collected along the last 70 ns in each trajectory, with each snapshot covering 200 ps. For the multi-trajectory runs, the mean position of individual trajectory was used to calculate the displacements before averaging when using equation (1). The fluctuation covariance matrix was used as input to construct a network following the method of Luthey-Schulten and co-workers [17]. Here every amino acid is defined as one node and every nucleotide is defined as two nodes with the nucleotide backbone as one and the base group as the other. An edge between two nodes is defined if they are not covalently bonded but are with heavy atoms closer than 4.5 Å over 75% of sampling time. The strength of the edge is defined as the absolute value of the covariance matrix element (C_{mn}) as computed in equation (1). The number of connected edges at each node is defined as the degree of the node. The correlation-weighted degree (CWD) is the normalized sum of all strengths of connected edges (apparently it is reduced to degree if equal weight is given to each edge). The relative correlation-weighted degree of the node is defined as the ratio of the total strength of all edges connected to the node over the total strengths of all nodes, indicating the importance of the node.

The definitions of the topological parameters from social network [49] and associated algorithms were utilized to describe and compare different networks. Central nodes are the most important nodes in network. Network centralization [50] indicates how centralized the central nodes are within a network: network with higher centralization is more centralized on a fewer central nodes. Network density quantifies the proportion of actual edges out of all possible edges, which indicates how easily the information can transfer in the network. Heterogeneity measures the uniformity of the network: higher heterogeneity indicates that the network is more likely consisted of different components. Hub node is the node which the ratio of CWDs for single node over CWDs of total nodes is higher than 1%.

As an extension of the method documented in Ref [17], network topological analysis and network plot were performed using Cytoscape 3.1.1 [51]. The Floyd-Warshall algorithm [34] was used to calculate the shortest path between any two nodes in the network. Edge betweenness is then defined as how many times this edge is passed by all the shortest paths. Thus the edge betweenness can be used to identify the importance of an edge in information flow. Therefore an edge is more important for information transfer if it has a higher betweenness [34]. In network visualization, edges are colored with their betweenness, thin green lines for lower betweenness and thick red lines for higher betweenness.

The networks were then clustered into communities with the Girvan-Newman algorithm [17] to summarize the difference between different networks. This algorithm is based on a process of progressive edge removal from the original network. At every step, betweenness of every edge is calculated, and the edge with the highest betweenness is removed. Once this edge is removed, the betweenness of every other edge is calculated and the edge of the highest betweenness is removed again. The process is repeated until no edge is left. Obviously important edges are removed first. The network can then be transformed into a hierarchical tree by backtracking the edge removal process. By inspecting the main branches of the hierarchical tree, functional communities can be identified. Nodes in the same

community are considered to be more closely correlated than those in different communities. Edges connecting different communities are considered as the most important paths for information flow [17].

Besides the general topology and community splitting, the network difference was also quantified with the proportion (P) of edge strengths (described above) between the protein and its target DNA out of the total edge strengths in the complex, as shown below:

$$P=\frac{\sum S_{\mathrm{btwn}}}{\sum S_{\mathrm{all}}} \quad (2)$$

where S_{btwn} is the strength of an edge between the protein and the DNA, and S_{all} is the strength of any edge within the whole complex. Because there are more nodes in ERGi than in ERGu, it may be confusing whether the proportion change is caused by the decrease between protein and DNA or by the increase of total edge strengths. To remove this uncertainty, proportions of edge strengths in all complexes were computed without considering nodes and edges that belong to the extra NID and CID in ERGi.

Other Computational Details

The last 50 ns of all trajectories were used to compute the average structures for the DNA binding domains in holo-ERGi and holo-ERGu. Tertiary contact assignment was handled with in-house software [40, 52, 53]. Average structures were reported as the snapshots closest (in term of main-chain RMSd) to the averaged main-chain structures after alignment using the PTRAJ program in AMBER12. SPARTA version 1.01 was used to predict chemical shift of nitrogen and hydrogen atoms [31]. These data can be compared with the N-H heteronuclear single-quantum coherence (HSQC) results in the previous experiments [19] as shown in Fig. 4. Linear regressions were performed using the R package [54].

In this study, the soft-core thermodynamic integration (TI) method [55, 56] was used to calculate the relative binding free energy difference between WT and mutants complexes, which can be compared directly with experiment. Taking WT of holo-ERGu or holo-ERGi as example, the three average WT structures from the three independent trajectories were used as the starting state. The corresponding Y354F mutant structures were used as the ending state. Coupling parameter λ was set to change with 0.05 intervals. Parallel version of PMEMD [55] was used to perform the soft-core TI calculation. For every λ window, the starting structure was first minimized with the steepest descent method for 10000 steps, and then followed by 25 ps heating and 100 ps equilibration. Production run lasted 1.5 ns for every λ. Supplementary Figure S1 shows the ΔDelta;G of Y354F of both ERGu and ERGi converges after 1.0 ns. Therefore, 1.5 ns simulations are sufficient for the equilibration of TI calculation.

ERG Protein Expression and Purification

Genes encoding ERGi (272-388) and ERGu (289-378) were cloned into the pET28a (Novagen) vector, with an N-terminal 6×His-tag. In order to confirm the allosteric pathway, a series of mutants were constructed by the whole-plasmid PCR and *Dpn*I digestion method,

and were verified by plasmid sequencing. The primers for all the constructs are listed in Table 5.

All ERG protein constructs were overexpression in *Escherichia coli* BL21 (DE3) cells (Novagen). Cells were cultured at 37°C in Luria Broth medium to an OD_{600} of 0.8–1.0, and were then induced for 20 hr at 16°C with 0.2 mM IPTG. Cells were harvested by centrifugation with 4000 rpm for 30 minutes at 4°C, and the cell pellet was resuspended in ice-cold binding buffer (25 mM MES, pH 6.0, 300 mM NaCl, and 20 mM imidazole). Resuspended cells were lysed by sonication, followed by centrifugation. The resulting supernatant of cell lysates was than purified on Ni^{2+}-NTA affinity chromatography (Qiagen) and washed thoroughly with 1 M KCl to remove contaminating DNA fragments [19]. Then the protein was further purified by Superdex75 gel filtration chromatography (GE Healthcare). The Superdex75 buffer contained 10 mM HEPES, pH 7.5 and 300 mM NaCl. Protein homogeneity was checked by SDS-PAGE followed by Coomassie-based staining (see Supplementary Fig. S2a). Peak fractions were combined with a final concentration of 6–7 mg/ml, flash-frozen in liquid nitrogen, and stored in −80°C until use.

Surface Plasmon Resonance (SPR)

1) ERG Protein Immobilization—All SPR experiments were performed at 25°C using a Biacore™ T200 instrument (GE Healthcare). All solutions and buffers were prepared with MilliQ water. The running buffer was HBS-EP buffer contained 10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, and 0.05% (v:v) surfactant P20 (GE Healthcare). All the ERG proteins were individually immobilized on CM5 sensor chips using a Biacore™ Amine Coupling Kit according to the manufacturer's instructions. Briefly, sensor chip surface was activated with two injections of N-hydroxysuccinimide (NHS, 0.1 M) and 1-ethyl-3-(3-dimethylaminopropyl)- carbodiimide (EDC, 0.4 M) (1:1) for 420 s at a flow rate of 10 μl/min. Each ERG protein, diluted to 20 μg/ml in 10 mM acetate (pH 4.0), was injected for 30 s at a flow rate of 10 μl/min. Excess free amine groups on sensor chip surface were blocked injecting 1 M ethanolamine-HCl (1.0 M, pH 8.5) 420 s at a flow rate of 10 μl/min. Typical immobilization level is ~2000 resonance units (RU).

2) Biacore Kinetic Binding Assays—Protein-bound DNA is a 12 bp double-strand oligonucleotide, with sequence 5'-GACC**GGAA**GTGG-3' as in MD simulations. It was synthesized and diluted in running buffer to final concentrations gradient of 200, 100, 50, 25, 12.5, 6.25, 3.125, 1.5625, 0.78125, 0.390625, 0.1853125 nM. Diluted samples were injected for 45 s at a flow rate of 30 μl/min over the derivatized surface, or an underivatized surface as a background control. Running buffer was then washed for 90 s at a flow rate of 30 μl/min, and finally the chip surface was regenerated by EP buffer for 30 s at a flow rate of 30 μl/min.

3) Curve fitting and modeling—Kinetic data and affinities were obtained from Biacore™ T200 Evaluation Software (version 2.0; GE Healthcare). The affinity evaluation fits a plot of response against concentration to the equation (3) [57].

$$R_{eq}=\frac{K_A C R_{max}}{K_A C+1}+\text{offset} \quad (3)$$

Here, K_A is the association constant, C is the DNA concentration (nM), R_{max} is the response at saturation, and offset is a constant term that gives the intercept of the fitted curve on the y-axis. Data was analyzed using a one-site binding model. The affinity is reported as a K_D value, which is the inverse of K_A. In addition, our SPR experiments were validated by utilizing the self-consistency tests for SPR reported previously [58]. The affinity is reported as a K_D value and regression curve of K_D for different systems are shown in Supplementary Fig. S2.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Prof. John H. Bushweller and his group at University of Virginia for their generous help in setting up the experiment and in providing the ERG plasmid. This work was supported by Center for HPC at Shanghai Jiao Tong University, by grants from the Ministry of Science and Technology of China (2012CB721003), the National High-Tech R&D Program of China (863 Program) (2014AA021502), the National Natural Science Foundation of China (J1210047 and 31620103901), Medical Engineering Cross Fund of Shanghai Jiao Tong University (YG2013MS68, YG2014MS47, and YG2015MS56), and National Institutes of Health/NIGMS (GM093040 & GM079383).

References

1. Dhand R, Hiles I, Panayotou G, Roche S, Fry MJ, Gout I, Totty NF, Truong O, Vicendo P, Yonezawa K, Yonezawa K, Kasuga M, Courtneidge SA. PI 3-kinase is a dual specificity enzyme: autoregulation by an intrinsic protein-serine kinase activity. EMBO J. 1994; 13:522–33. [PubMed: 8313897]
2. Sharrocks AD. The ETS-domain transcription factor family. Nat Rev Mol Cell Biol. 2001; 2:827–837. [PubMed: 11715049]
3. Favelyukis S, Till JH, Hubbard SR, Miller WT. Structure and autoregulation of the insulin-like growth factor 1 receptor kinase. Nat Struct Mol Biol. 2001; 8:1058–1063.
4. Brinkman AB, Ettema TJ, de Vos WM, van der Oost J. The Lrp family of transcriptional regulators. Mol Microbiol. 2003; 48:287–294. [PubMed: 12675791]
5. Herdegen T, Leah J. Inducible and constitutive transcription factors in the mammalian nervous system: control of gene expression by Jun, Fos and Krox, and CREB/ATF proteins. Brain Res Rev. 1998; 28:370–490. [PubMed: 9858769]
6. Tanji H, Ohto U, Motoi Y, Shibata T, Miyake K, Shimizu T. Autoinhibition and relief mechanism by the proteolytic processing of Toll-like receptor 8. Proc Natl Acad Sci U S A. 2016; 113:3012–3017. [PubMed: 26929371]
7. Laughlin JD, Nwachukwu JC, Figuera-Losada M, Cherry L, Nettles KW, LoGrasso PV. Structural mechanisms of allostery and autoinhibition in JNK family kinases. Structure. 2012; 20:2174–84. [PubMed: 23142346]
8. Viaud J, Peterson JR. An allosteric kinase inhibitor binds the p21-activated kinase autoregulatory domain covalently. Mol Cancer Ther. 2009; 8:2559–2565. [PubMed: 19723886]
9. Johnson PF, McKnight SL. Eukaryotic transcriptional regulatory proteins. Annu Rev Biochem. 1989; 58:799–839. [PubMed: 2673023]

10. Kobe B, Jennings IG, House CM, Michell BJ, Goodwill KE, Santarsiero BD, Stevens RC, Cotton RG, Kemp BE. Structural basis of autoregulation of phenylalanine hydroxylase. Nat Struct Mol Biol. 1999; 6:442–448.
11. Domitrovic T, Movahed N, Bothner B, Matsui T, Wang Q, Doerschuk PC, Johnson JE. Virus assembly and maturation: auto-regulation through allosteric molecular switches. J Mol Biol. 2013; 425:1488–96. [PubMed: 23485419]
12. Peterson JR, Golemis EA. Autoinhibited proteins as promising drug targets. J Cell Biochem. 2004; 93:68–73. [PubMed: 15352163]
13. Agafonov RV, Wilson C, Otten R, Buosi V, Kern D. Energetic dissection of Gleevec's selectivity toward human tyrosine kinases. Nat Struct Mol Biol. 2014; 21:848–53. [PubMed: 25218445]
14. Chen YN, LaMarche MJ, Chan HM, Fekkes P, Garcia-Fortanet J, Acker MG, Antonakos B, Chen CH, Chen Z, Cooke VG, Dobson JR, Deng Z, Fei F, Firestone B, Fodor M, Fridrich C, Gao H, Grunenfelder D, Hao HX, Jacob J, Ho S, Hsiao K, Kang ZB, Karki R, Kato M, Larrow J, La Bonte LR, Lenoir F, Liu G, Liu S, Majumdar D, Meyer MJ, Palermo M, Perez L, Pu M, Price E, Quinn C, Shakya S, Shultz MD, Slisz J, Venkatesan K, Wang P, Warmuth M, Williams S, Yang G, Yuan J, Zhang JH, Zhu P, Ramsey T, Keen NJ, Sellers WR, Stams T, Fortin PD. Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. Nature. 2016; 535:148–52. [PubMed: 27362227]
15. Hengge R. Principles of c-di-GMP signalling in bacteria. Nat Rev Microbiol. 2009; 7:263–273. [PubMed: 19287449]
16. Changeux JP. 50 years of allosteric interactions: the twists and turns of the models. Nat Rev Mol Cell Biol. 2013; 14:819–29. [PubMed: 24150612]
17. Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA: protein complexes. Proc Natl Acad Sci U S A. 2009; 106:6620–6625. [PubMed: 19351898]
18. Palazzesi F, Barducci A, Tollinger M, Parrinello M. The allosteric communication pathways in KIX domain of CBP. Proc Natl Acad Sci U S A. 2013; 110:14237–14242. [PubMed: 23940332]
19. Regan MC, Horanyi PS, Pryor EE, Sarver JL, Cafiso DS, Bushweller JH. Structural and dynamic studies of the transcription factor ERG reveal DNA binding is allosterically autoinhibited. Proc Natl Acad Sci U S A. 2013; 110:13374–13379. [PubMed: 23898196]
20. Degnan BM, Degnan SM, Naganuma T, Morse DE. The ets multigene family is conserved throughout the Metazoa. Nucleic Acids Res. 1993; 21:3479–3484. [PubMed: 8346026]
21. Lelièvre E, Lionneton F, Soncin F, Vandenbunder B. The Ets family contains transcriptional activators and repressors involved in angiogenesis. Int J Biochem Cell B. 2001; 33:391–407.
22. Tomlins SA, Mehra R, Rhodes DR, Smith LR, Roulston D, Helgeson BE, Cao X, Wei JT, Rubin MA, Shah RB. TMPRSS2: ETV4 gene fusions define a third molecular subtype of prostate cancer. Cancer Res. 2006; 66:3396–3400. [PubMed: 16585160]
23. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science. 2005; 310:644–648. [PubMed: 16254181]
24. Hsu T, Trojanowska M, Watson DK. Ets proteins in biological control and cancer. J Cell Biochem. 2004; 91:896–903. [PubMed: 15034925]
25. Codrington R, Pannell R, Forster A, Drynan LF, Daser A, Lobato N, Metzler M, Rabbitts TH. The Ews-ERG fusion protein can initiate neoplasia from lineage-committed haematopoietic cells. PLoS Biol. 2005; 3:e242. [PubMed: 15974803]
26. Donaldson LW, Mcintosh LP, Skalicky JJ, Petersen JM, Graves BJ. Structural coupling of the inhibitory regions flanking the ETS domain of murine Ets-1. Protein Sci. 1996; 5:296–309. [PubMed: 8745408]
27. Lee GM, Donaldson LW, Pufall MA, Kang HS, Pot I, Graves BJ, McIntosh LP. The structural and dynamic basis of Ets-1 DNA binding autoinhibition. J Biol Chem. 2005; 280:7088–7099. [PubMed: 15591056]
28. Lee GM, Pufall MA, Meeker CA, Kang HS, Graves BJ, McIntosh LP. The affinity of Ets-1 for DNA is modulated by phosphorylation through transient interactions of an unstructured region. J Mol Biol. 2008; 382:1014–1030. [PubMed: 18692067]

29. Xu L, Ye W, Jiang C, Yang J, Zhang J, Feng Y, Luo R, Chen HF. Recognition Mechanism between Lac Repressor and DNA with Correlation Network Analysis. J Phys Chem B. 2015; 119:2844–2856. [PubMed: 25633018]
30. Zhang J, Luo H, Liu H, Ye W, Luo R, Chen H-F. Synergistic Modification Induced Specific Recognition between Histone and TRIM24 via Fluctuation Correlation Network Analysis. Sci Rep. 2016; 6
31. Shen Y, Bax A. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J Biomol NMR. 2007; 38:289–302. [PubMed: 17610132]
32. Coyne HJ III, De S, Okon M, Green SM, Bhachech N, Graves BJ, McIntosh LP. Autoinhibition of ETV6 (TEL) DNA binding: appended helices sterically block the ETS domain. J Mol Biol. 2012; 421:67–84. [PubMed: 22584210]
33. De S, Chan AC, Coyne HJ III, Bhachech N, Hermsdorf U, Okon M, Murphy ME, Graves BJ, McIntosh LP. Steric Mechanism of Auto-Inhibitory Regulation of Specific and Non-Specific DNA Binding by the ETS Transcriptional Repressor ETV6. J Mol Biol. 2014; 426:1390–1406. [PubMed: 24333486]
34. Floyd RW. Algorithm 97: shortest path. Commun ACM. 1962; 5:345.
35. The PyMOL Molecular Graphics System, version 1.7.0.0. Schrodinger, LLC; 2014.
36. SYBYL-X Suite, version 2.1.1. Certara; Princeton, NJ: 2012.
37. Case DA, Darden TA, Cheatham ITE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Hayik S, Roitberg A, Seabra G, Swails J, Götz AW, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wolf RM, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Salomon-Ferrer R, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA. AMBER 12. 2012
38. Darden T, York D, Pedersen L. Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems. J Chem Phys. 1993; 98:10089–10092.
39. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins: Struct, Funct Bioinf. 2010; 78:1950–1958.
40. Wang W, Ye W, Jiang C, Luo R, Chen HF. New force field on modeling intrinsically disordered proteins. Chem Biol Drug Des. 2014; 84:253–269. [PubMed: 24589355]
41. Ye W, Ji D, Wang W, Luo R, Chen HF. Test and Evaluation of ff99IDPs Force Field for Intrinsically Disordered Proteins. J Chem Inf Model. 2015; 55:1021–1029. [PubMed: 25919886]
42. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977; 23:327–341.
43. Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. J Chem Theory Comput. 2012; 8:1542–1555. [PubMed: 22582031]
44. Götz AW, Salomon-Ferrer R, Poole D, Grand S, Walker R. Routine microsecond molecular dynamics simulations with AMBER. Part II: particle mesh Ewald. J Chem Theory Comput. 2013; 9:3878–3888. [PubMed: 26592383]
45. Tai K, Shen T, Börjesson U, Philippopoulos M, McCammon JA. Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. Biophys J. 2001; 81:715–724. [PubMed: 11463620]
46. Young MA, Gonfloni S, Superti-Furga G, Roux B, Kuriyan J. Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. Cell. 2001; 105:115–126. [PubMed: 11301007]
47. Hünenberger P, Mark A, Van Gunsteren W. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. J Mol Biol. 1995; 252:492–503. [PubMed: 7563068]
48. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins: Struct, Funct Bioinf. 1991; 11:205–217.

49. Sparrowe RT, Liden RC, Wayne SJ, Kraimer ML. Social networks and the performance of individuals and groups. Acad Manag J. 2001; 44:316–325.
50. Dong J, Horvath S. Understanding network concepts in modules. BMC Syst Biol. 2007; 1:24. [PubMed: 17547772]
51. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–2504. [PubMed: 14597658]
52. Chen HF, Luo R. Binding induced folding in p53-MDM2 complex. J Am Chem Soc. 2007; 129:2930–2937. [PubMed: 17302414]
53. Ye W, Yang J, Yu Q, Wang W, Hancy J, Luo R, Chen HF. Kink turn sRNA folding upon L7Ae binding using molecular dynamics simulations. Phys Chem Chem Phys. 2013; 15:18510–18522. [PubMed: 24072031]
54. Ihaka R, Gentleman R. R: a language for data analysis and graphics. J Comput Graph Stat. 1996; 5:299–314.
55. Kaus JW, Pierce LT, Walker RC, McCammon JA. Improving the efficiency of free energy calculations in the AMBER molecular dynamics package. J Chem Theory Comput. 2013; 9:4131–4139.
56. Case, D., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D., Cheatham, T., Iii, Darden, T., Duke, R., Gohlke, H. AMBER 14. University of California; San Francisco: 2014.
57. Schuck P, Minton AP. Kinetic analysis of biosensor data: elementary tests for self-consistency. Trends Biochem Sci. 1996; 21:458–460. [PubMed: 9009825]
58. Cornish-Bowden A. Detection of errors of interpretation in experiments in enzyme kinetics. Methods. 2001; 24:181–190. [PubMed: 11384193]

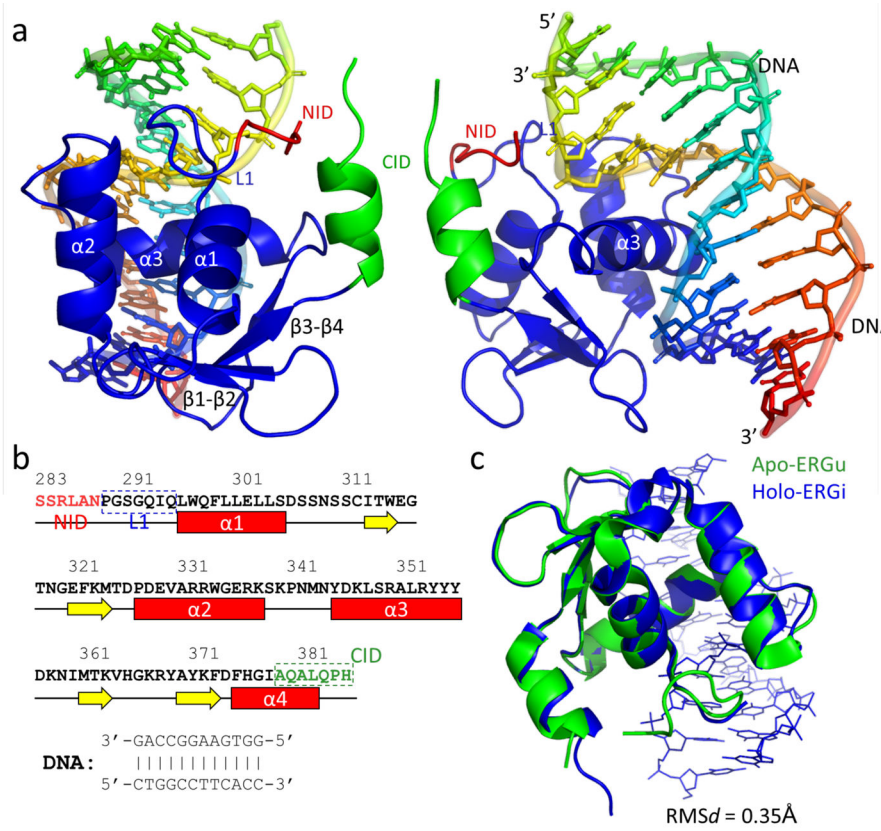


Figure 1.

Initial structure and sequence information of autoinhibited ERG/DNA complex. (a) Cartoon representations with two viewing angles. The conserved core domain (blue) binds with the DNA GGA motif at the interface of helix α3 and the major groove. NID (red) and CID (green), which locate at the two termini of the core domain, are in contact with each other tightly. (b) Sequence and secondary structure of the autoinhibited ERG/DNA complex. (c) Structural alignment between autoinhibited (blue) and uninhibited (green), ERGs showing high structural similarity.

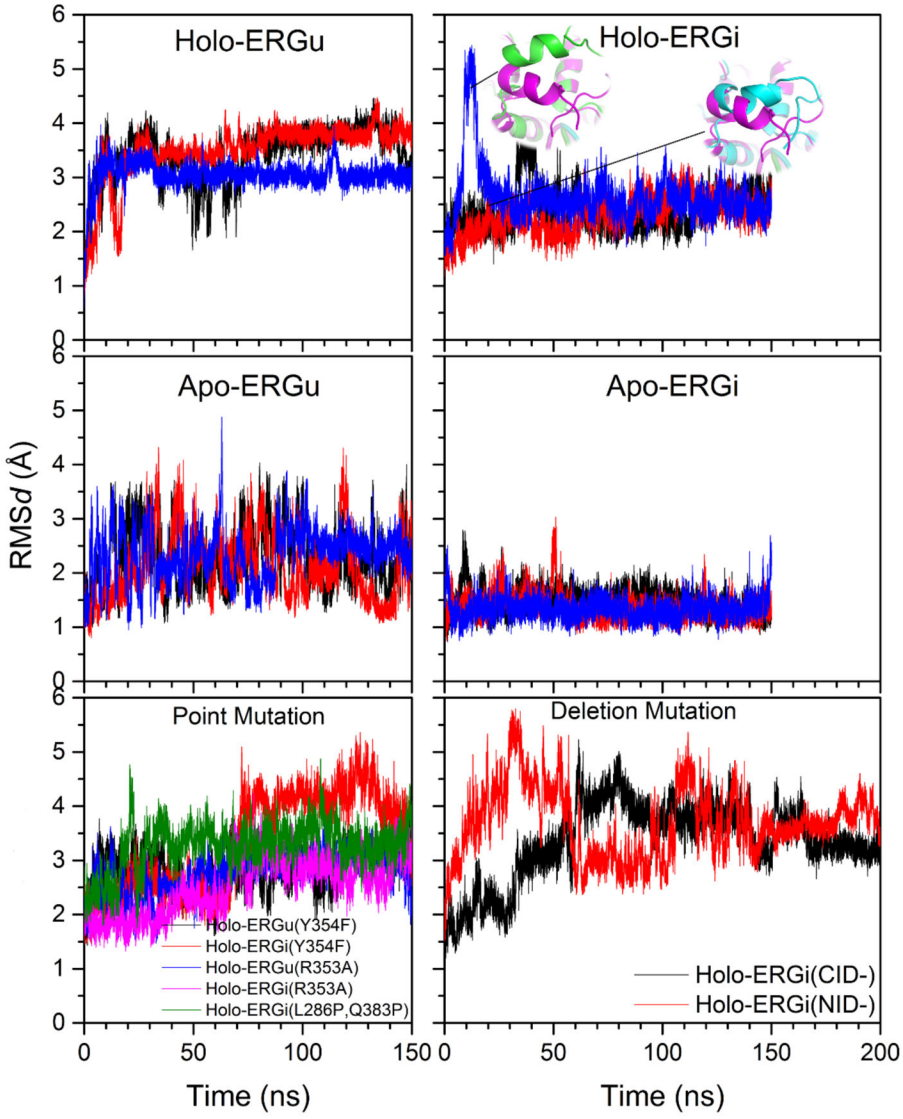


Figure 2.

RMS*d* for all simulated trajectories. There are three trajectories in the top four plots. For mutant systems, only one trajectory was simulated. Worth noting is the blue trajectory for holo-ERGi where a large fluctuation is observed at the beginning of the simulation. This is due to the large oscillation of NID and CID, as illustrated in the structural alignments (insert in the top right plot) among the initial structure (magenta), the snapshot with highest RMS*d* (at 15ns, green), and the snapshot that goes back to the equilibrium (at 30ns, blue), which shows high structural consistency in the core domain and variation mainly in the flexible NID and CID.

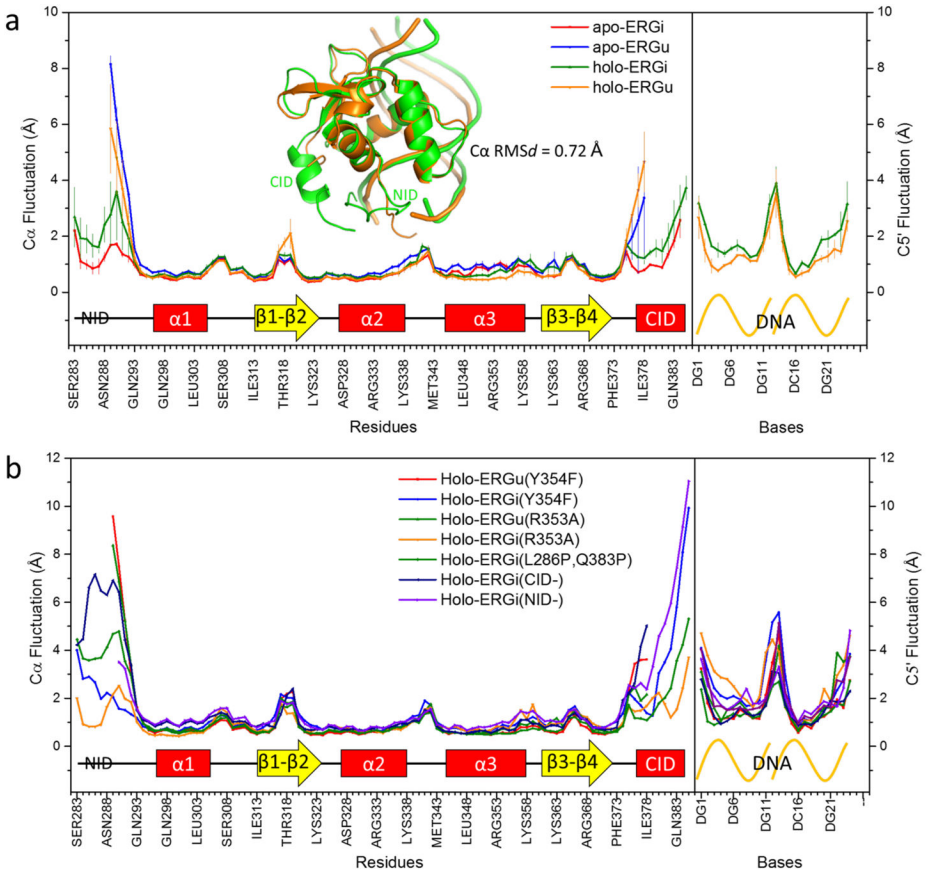


Figure 3.

RMS*f* for all the simulation trajectories with secondary structures labeled. (a) Wild-type trajectories. Mean structures of holo-ERGi (green) and holo-ERGu (orange) are also shown and aligned with Cα atoms, RMS*d* is 0.72Å.; (b) Mutant trajectories.

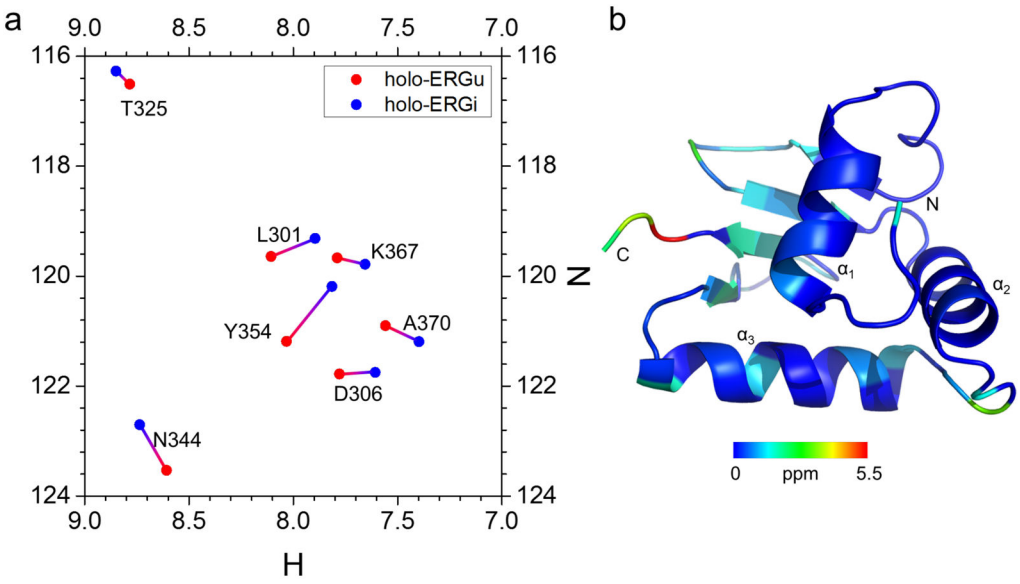


Figure 4.

Comparison of chemical shift perturbation between holo-ERGi and holo-ERGu with previous experiment [19]. Chemical shift data of atoms N and H for all residues (except proline) were calculated using SPARTA 1.01 [31]. (a) N-H HSQC spectra between holo-ERGu (red) and holo-ERGi (blue), which can be compared with Fig. 3E in Regan et al. [19]. (b) Chemical shift perturbations mapped on the cartoon representation of protein part of holo-ERGu, with color scale labeled below. This can be compared with Supplementary Fig. S2 in Regan et al.

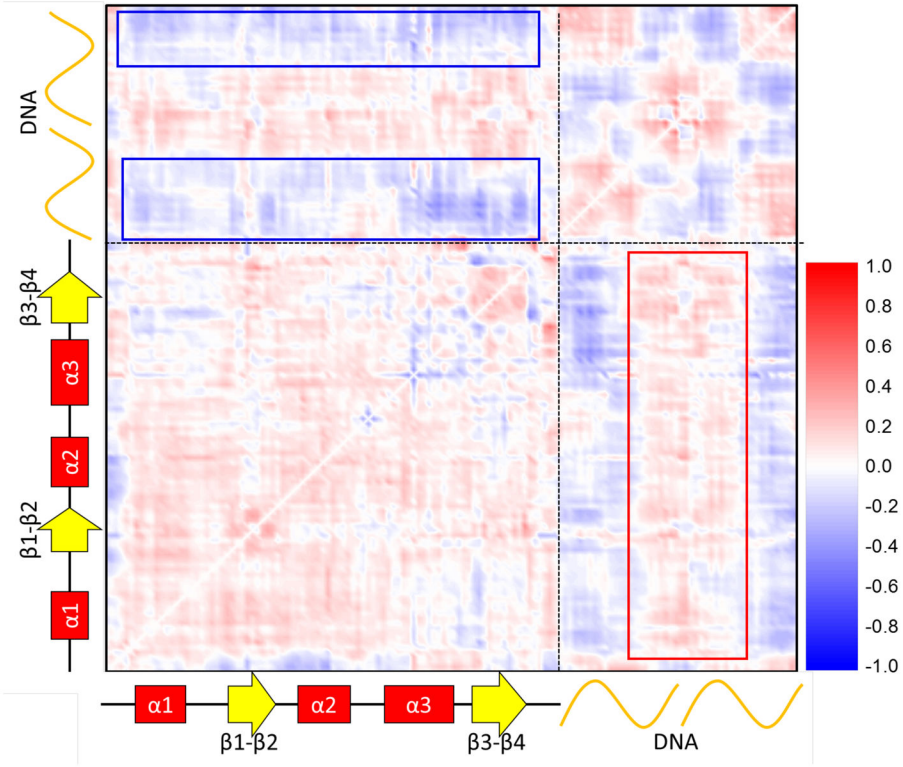


Figure 5.

Averaged differences in inter-residue correlation coefficients between holo-ERGi and holo-ERGu (C_{ERGi} minus C_{ERGu}), secondary structures were labeled. Two dashed lines split the contour into four regions: top-left and bottom-right for protein-DNA correlations, bottom-left for intramolecular correlations within protein, and top-right for the intramolecular correlations within DNA. Two blue boxes in the top-left region indicates that correlations are weaker between protein and DNA in the presence of NID and CID, i.e. protein and DNA are less correlated under autoinhibited state; The red box on the bottom-right shows the correlations within protein are higher under autoinhibited state in the presence of NID and CID.

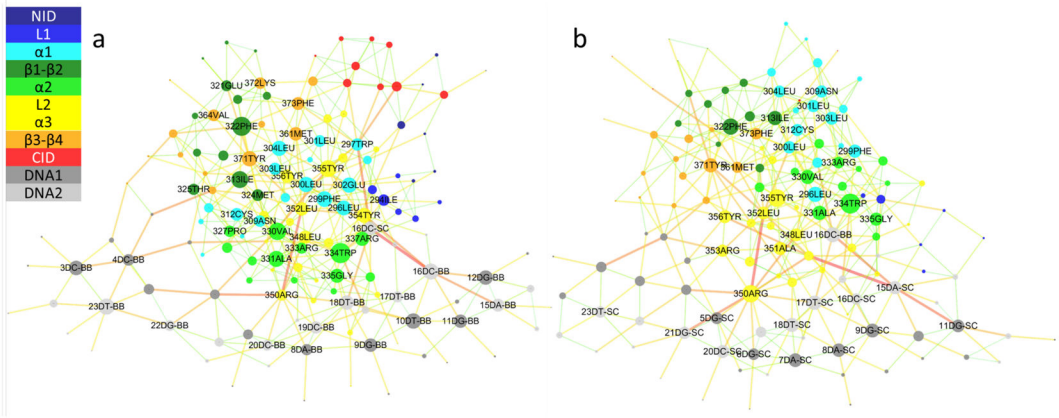


Figure 6.

Dynamical networks on fluctuation correlation of (a) holo-ERGi and (b) holo-ERGu. Nodes are drawn to sizes based on the actual values for the ratio of CWDs and are colored according to their structural domains. See labels in the coloring legend. Hub nodes are labeled with the residues/subunits.

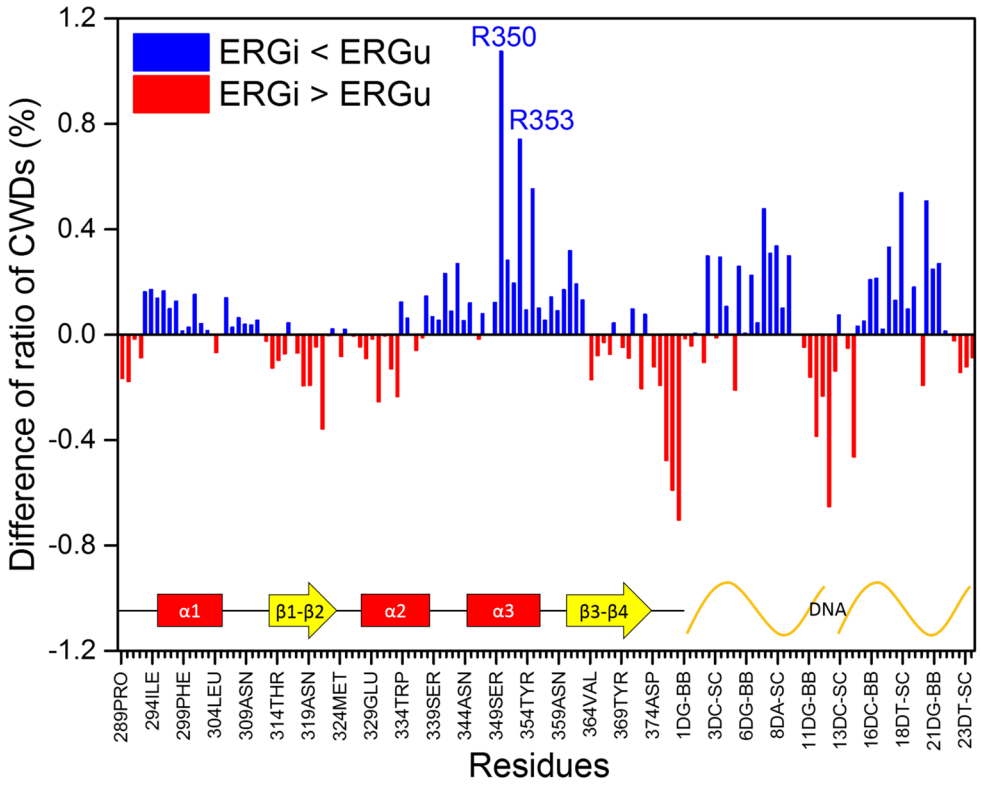


Figure 7.

Difference in the ratio of CWDs between holo-ERGi and holo-ERGu, with red bars as ERGi > ERGu, and blue bars as ERGi < ERGu. The DNA binding interface is located at the α3 helix.

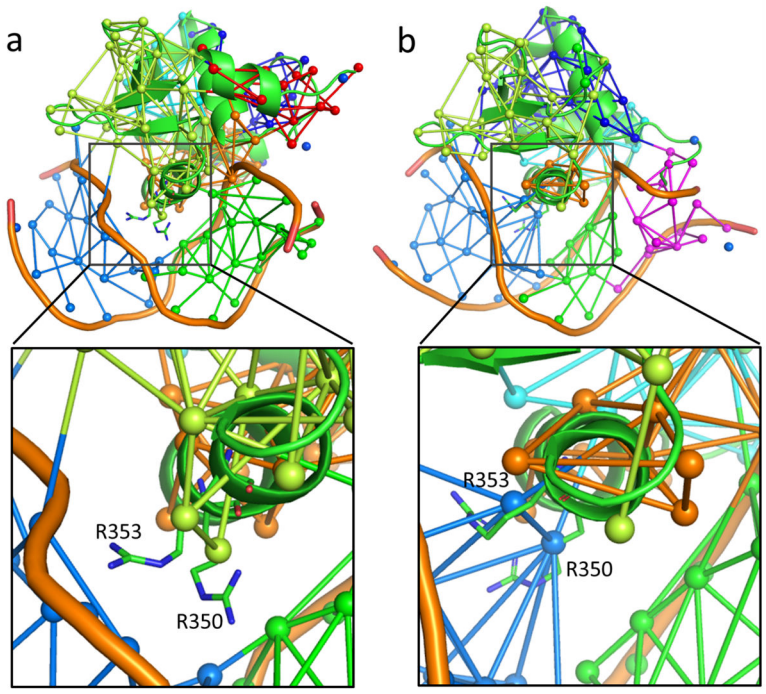


Figure 8.

Network communities on (a) holo-ERGi and (b) holo-ERGu. Communities were split using the Girvan-Newman algorithm based on the edges and correlations between nodes in the correlation networks. The corresponding zoom-in details at the binding interface are shown at the bottom. Edges can be found between communities of α3 and DNA in holo-ERGu. However, α3 and DNA are not closely connected in holo-ERGi. In each network, different colors show different communities.

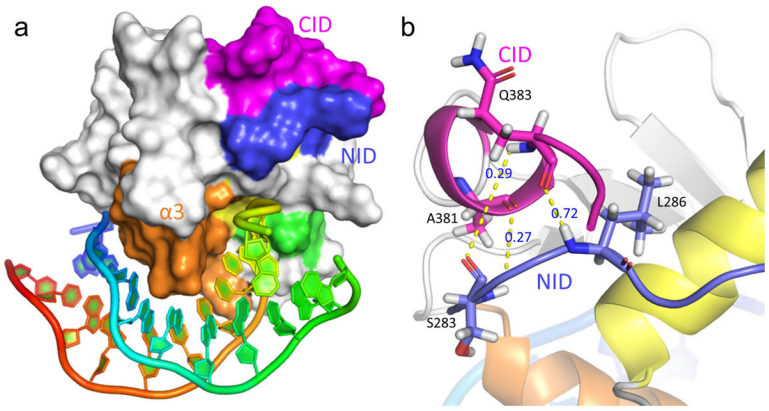


Figure 9.

Potential allostery induced from the NID-CID binding interface. (a) NID (blue) and CID (magenta) are in contact closely, but neither of them contact with DNA directly. In contrast, helix α3 (orange), far away from NID and CID, binds with the major groove of DNA. (b) Hydrogen-bond network between NID and CID in holo-ERGi. Numbers in dark blue represent the average population for each hydrogen bond. The residues involved in the hydrogen-bond network are thus considered as key residues in the binding between NID and CID.

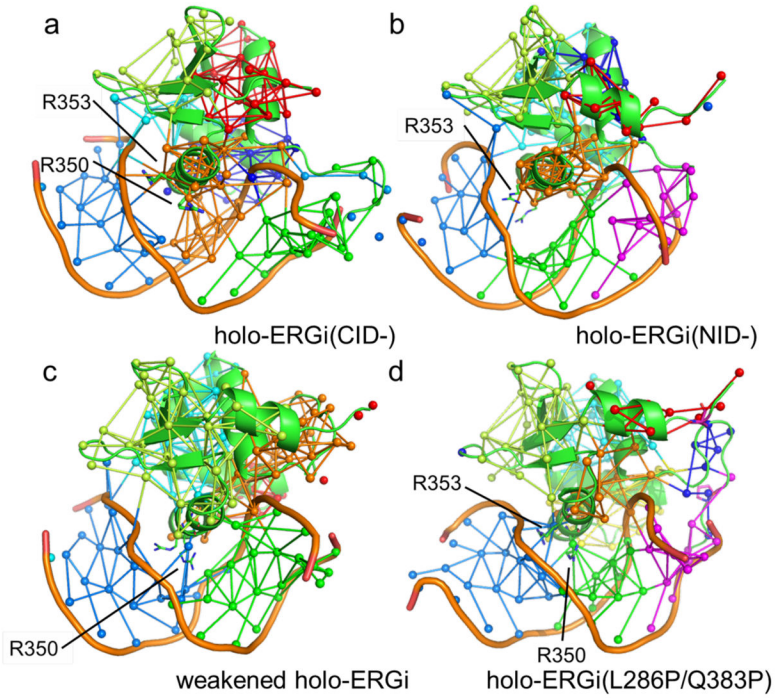


Figure 10.

Network communities on (a) holo-ERGi(CID-), (b) holo-ERGi(NID-), (c) connectivity-weakened holo-ERGi, and (d) holo-ERGi(L286P/Q383P). All these deleted/mutant complexes show uninhibited-like community splitting, which is distinctly different from the original complex of autoinhibited holo-ERGi, confirming the binding between NID and CID is key to the changing of correlation network in the complex.

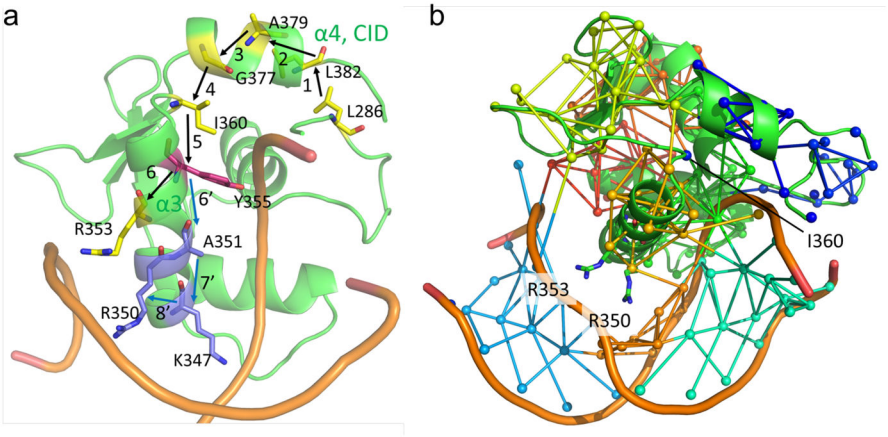


Figure 11.

Allosteric autoinhibition pathways and community network analysis. (a) Proposed allosteric autoinhibition pathways in holo-ERGi. (b) Community network after weakening the connectivity. CWDs of all I360's edges are set to zero in the network perturbation analysis. An uninhibited-like community-splitting outcome becomes apparent, with R350 connecting communities of α3 and DNA.

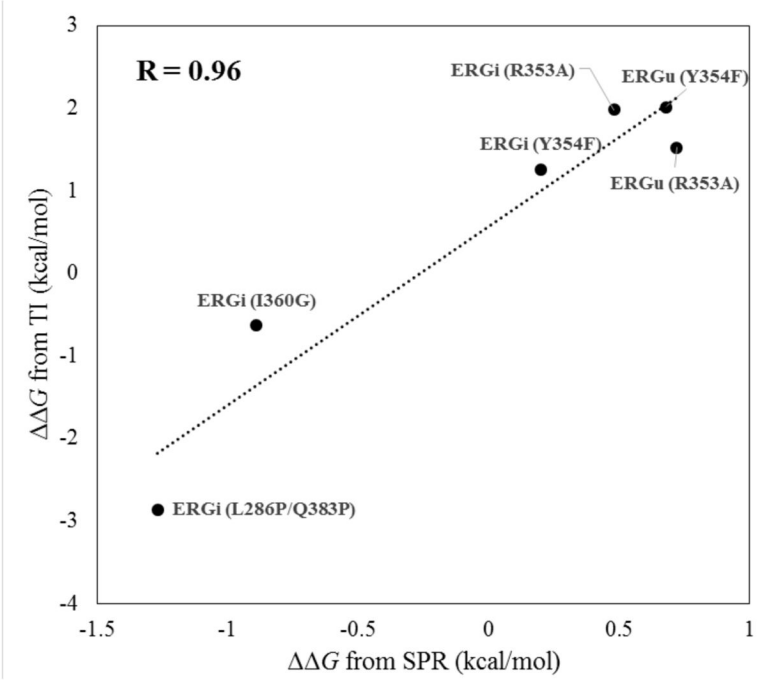


Figure 12.
Linear correlation between Gs derived from SPR measurements and TI simulations.

Table 1

Network topological characteristics of holo-ERGi and holo-ERGu.

Characteristics	Holo-ERGi	Holo-ERGi (ΔNID&CID)	* Holo-ERGu
Network centralization	0.060	0.067	0.065
Network density	0.036	0.041	0.041
Heterogeneity	0.542	0.546	0.546
Average Num. of neighbors	5.293	5.364	5.507

* Only regions without NID and CID in holo-ERGi were analyzed.

Table 2

Relative binding free energies (with respect to WT) for ERGi and ERGu.

System	Δ*G*(kcal/mol) from MMGBSA	ΔΔ*G* (kcal/mol) from TI	K_D (nM) from SPR	Δ Δ*G* (kcal/mol) from SPR
ERGu	−127.13±14.28	0	7.22±1.40	0
ERGu(R353A)	−93.41±15.52	1.52±2.27	24.84±8.40	+0.72
ERGu(Y354F)	−102.16±15.79	2.01±0.74	22.62±7.00	+0.68
ERGi	−107.11±13.88	0	37.18±1.70	0
ERGi(R353A)	−65.79±12.55	1.98±1.30	83.18±5.60	+0.48
ERGi(Y354F)	−70.20±14.92	1.26±1.04	52.44±22.0	+0.20
ERGi(L286P/Q383P)	−118.57±14.91	−2.87±1.54	4.33±0.73	−1.27
ERGi(I360G)	−110.68±15.81	−0.63±0.56	8.29±0.19	−0.89

Table 3

Raw correlation coefficients between R350, R353, Y354 and DNA in the holo-ERGi and holo-ERGu.

holo-ERGi		holo-ERGu	
Connections	**Strength**	**Connections**	**Strength**
R350 - C4_SC	0.1628	R350 - C4_SC	0.3125
R350 - G5_SC	0.0868	R350 - G5_SC	0.3862
R350 - G6_SC	0.3462	R350 - G6_SC	0.2979
R350 - A7_SC	0.2884	R350 - A7_SC	0.1222
R350 - T17_SC	0.2645	R350 - T17_SC	0.2881
R350 - T18_SC	0.2821	R350 - T18_SC	0.3517
R350 - C19_SC	0.1645	R350 - C19_SC	0.3560
R350 - C20_SC	-0.0682	R350 - C20_SC	0.3599
R350 - G21_SC	-0.0138	R350 - G21_SC	0.2440
R353 - C3_BB	−0.038	R353 - C3_BB	0.3378
R353 - C4_BB	0.0948	R353 - C4_BB	0.4924
R353 - C4_SC	0.1461	R353 - C4_SC	0.3384
Y354 - C14_BB	0.1875	Y354 - C14_BB	0.0315
Y354 - A15_BB	0.5324	Y354 - A15_BB	0.3078
Y354 - A15_SC	0.4239	Y354 - A15_SC	0.3158
Y354 - C16_SC	0.3523	Y354 - C16_SC	0.3246

BB represents backbone. SC represents sidechain. The highlighted entries indicate that these pairs of nodes have no edges in the networks.

Table 4

Simulation conditions for all systems.

System	Temperature (K)	Time (ns)	Traj. Num.	Ions	Waters
Apo-ERGi		150	3	4 Cl^-	4331
Apo-ERGu		150	3	3 Cl^-	3564
Holo-ERGi		150	3	18 Na^+	11355
Holo-ERGu		150	3	15 Na^+	9603
Holo-ERGi(Y354F)		150	1	18 Na^+	6662
Holo-ERGi(R353A)	298	150	1	19 Na^+	6556
Holo-ERGu(Y354F)		150	1	19 Na^+	5776
Holo-ERGu(R353A)		150	1	18 Na^+	5233
Holo-ERGi(L286P/Q383P)		150	1	18 Na^+	5694
Holo-ERGi(CID-)		200	1	18 Na^+	5887
Holo-ERGi(NID-)		200	1	19 Na^+	6190

Table 5

Primers used in PCR.

Primer Name	Sequence (5’ to 3’)
5′ ERGi	GGAATTCCATATGCAGTTAGATCCTTATCAG
3′ ERGi	CCGCTCGAGTTACTCCGGGGGGTGGGGCTG
5′ ERGu	GGAATTCCATATGCCAGGCAGTGGCCAGATC
3′ ERGu	CCGCTCGAGTTAGATCCCGTGGAAGTCGAA
5′ R353A	GCCCTCgctTACTACTATGACAAGAACATCATGACCAA
3′ R353A	TAGTAGTAagcGAGGGCGCGGCTGAGCTTATC
5′ Y354F	CCTCCGTttcTACTATGACAAGAACATCATGACCAAGG
3′ Y354F	CATAGTAgaaACGGAGGGCGCGGCTGAGCTTA
5′ ERGi I360G	CAAGAACggcATGACCAAGGTCCATGGGAAGC
3′ ERGi I360G	TGGTCATgccGTTCTTGTCATAGTAGTAACGGAGGG
5′ ERGi L286P	TAGCCGCcctGCAAATCCAGGCAGTGGCC
3′ ERGi L286P	GATTTGCaggGCGGCTACTTGTTGGTCCAA
5′ ERGi Q383P	AGGCCCTCccgCCCCACCCCCCGGAGCTC
3′ ERGi Q383P	GTGGGGcggGAGGGCCTGGGCGATCCC