Autoinhibitory mechanisms of ERG studied by molecular dynamics simulations

Yan Lu and Freddie R. Salsbury, Jr.
Department of Physics, Wake Forest University, Winston-Salem, NC 27106, USA
(Received 2 December 2014; accepted 14 January 2015; published online 22 January 2015)

ERG, an ETS-family transcription factor, acts as a regulator of differentiation of early hematopoietic cells. It contains an autoinhibitory domain, which negatively regulates DNA-binding. The mechanism of autoinhibition is still illusive. To understand the mechanism, we study the dynamical properties of ERG protein by molecular dynamics simulations. These simulations suggest that DNA binding autoinhibition associates with the internal dynamics of ERG. Specifically, we find that (1), The N-C terminal correlation in the inhibited ERG is larger than that in uninhibited ERG that contributes to the autoinhibition of DNA-binding. (2), DNA-binding changes the property of the N-C terminal correlation from being anti-correlated to correlated, that is, changing the relative direction of the correlated motions and (3), For the Ets-domain specifically, the inhibited and uninhibited forms exhibit essentially the same dynamics, but the binding of the DNA decreases the fluctuation of the Ets-domain. We also find from PCA analysis that the three systems, even with quite different dynamics, do have highly similar free energy surfaces, indicating that they share similar conformations.

I. INTRODUCTION

Transcription factor binding is an important mechanism that regulates the gene expression which determines the cellular behavior.1 ERG, an ETS-family transcription factor, is a nuclear protein that binds purine-rich sequences of DNA.2,3 ERG may act as a regulator of differentiation of early hematopoietic cells,4 and is overexpressed in prostate tumors because of chromosomal rearrangement.5 Besides the DNA-binding domain, ERG contains an autoinhibitory domain, which negatively regulates DNA-binding. As a result, DNA-binding activity would be enhanced without this domain, which may result in a highly specific disease phenotype that involves gain of function rather than loss of function.6 Many members of ETS family contain autoinhibitory domain, but the inhibition mechanism may differ amongst different family members. It is proposed that autoinhibition can help distinguish among these related proteins.6

Autoinhibitory elements were first discovered in Ets-1 by the observation that deletion of regions flanking the ETS domain will enhance DNA-binding activity.7-9 Quantitative analyses showed that deletion of either the N-terminal or C-terminal flanking regions can strongly activate DNA binding, suggesting that two regions function together to mediate autoinhibition.10 More specifically, the core ETS domain contains three α-helices on a four-stranded anti-parallel β-sheet. Helices H1-1, H1-2, H4 and H5 inhibit DNA binding of Ets-1. NMR studies detected structural coupling between inhibitory helices, as well as a connection between the inhibitory elements and helix H1 of the ETS domain.11 Protease sensitivity measurements and CD spectroscopy analysis of alpha helicity showed that helix H1-1, which is distal from the recognition helix H3, unfolds upon DNA binding as part of the antoinhibition mechanisms.12 A second ETS member, ETV6, a transcriptional repressor,13 has also been found recently to be autoinhibited by Helix H5 in C inhibitory domain.
which would unfold when binding to DNA.\textsuperscript{14,15} DNA-binding autoinhibition has also been reported for other ETS factors. For example, members of the PEA3 subfamily have inhibitory sequences, but do not appear to be helical as determined by proline-scanning mutagenesis.\textsuperscript{16,17} As for ESE and TCF subfamilies, inhibition of DNA binding is mediated in complex manners.\textsuperscript{18–20}

Autoinhibitory regions of ERG were recently identified by Regan et. al.\textsuperscript{21} They have solved the crystal structures of uninhibited, auto-inhibited, and DNA-bound ERG. The NMR-based measurements of backbone dynamics showed that uninhibited ERG undergoes substantial dynamics and they proposed that the allosteric basis of ERG autoinhibition is mediated predominantly by the regulation of Ets-domain dynamics.\textsuperscript{21} In order to have a detailed insight into how the dynamics regulate the DNA binding, we performed in this article molecular dynamical simulations for the three different configurations. Due to their large size and long time scale, all-atom molecular dynamics is time consuming. Nevertheless, there is a useful coarse-grained model (one-particle-per-residue) called Cafemol. It has been successively used to study the dynamics of proteins and protein-DNA interactions.\textsuperscript{22–26} It has also been shown that it is a good model to study the correlated motions of proteins.\textsuperscript{27} Here we use Cafemol to study the inhibition mechanisms of the ERG protein in terms of fluctuations and correlated motions.

II. MODELS AND METHODS

A. Systems

The crystal structures of uninhibited, inhibited and DNA-binding ERG are available from the Protein Data Bank and were used as references (PDBID: 4IRG, 4IRH and 4IRI\textsuperscript{21} respectively). For clarity, We refer to them as ERGu, ERGi, ERGi:DNA complex. These three constructs spanning residues 286-383, 283-384, 292-385 of ERG respectively, containing helix 1 to 4 but none of H1-1 or H2-2. The DNA sequence is the consensus sequence of ACCGGAAGT.\textsuperscript{21} Only the shared residues are used for analysis.

B. Molecular dynamics

Langevin dynamics simulations were performed by using the Cafemol method with software version 2.1.\textsuperscript{28} This methodology implements multiple coarse-grained interaction models together in order to allow for the study of proteins, DNA, RNA and their interactions. These models are all based on different variants of off-lattice models. In this study, there are local interactions within the protein, non-local interactions within the protein, local interactions within the DNA, non-local interactions within the DNA, and non-local interaction between the DNA and the protein. Non-local and local is used in the same sense as all-atom molecular dynamics simulations, in that non-local interactions are potentially long-range interactions, whereas local interactions define the local covalent structure. For the protein, the off-lattice Go model is used which contains three local and two non-local terms with single-residue beads located at the alpha carbon. The local terms are a harmonic “virtual” bond term, virtual since coarse-grained beads are not actually bonded in the molecular structure, but are constrained, a harmonic angle term, and sinusoidal dihedral terms. The non-local terms are a Go-type interaction between pairs of beads that are in contact, yet not involved in a local interaction in the native structure, and repulsive interactions, i.e., the repulsive portion of a Lennard-Jones interaction, between pairs of beads that are not in contact in the native structure, and also not involved in a local interaction. For the DNA, the potential is more complicated and the coarse-graining requires a three-site model for the sugars, phosphates and bases. This complexity is necessary to capture the ionic strength-dependence of melting temperature, persistence length and heat capacity of double-stranded DNA. The local interactions include the same types as for the protein, albeit with different parameters for bond, angle and dihedral potentials. The non-local interactions terms model base-stacking, Watson-Crick base-pairing, excluded volume, and electrostatic interactions as well as solvation model for the sugars. The base-stacking interaction is modeled as a Lennard-Jones interaction between bases within a single strand of DNA with a 0.9 nm cutoff. The base-pairing interaction is modeled as a Go-type interaction between Watson-Crick paired bases on different strands. The DNA electrostatics are modeled solely as Debye-Huckel interactions between phosphate beads that are not bonded. For the protein-DNA interactions, electrostatic interactions
were used to model non-specific interactions between the protein and the DNA, in which only the phosphate beads of the DNA, and the charged amino acids within the protein are considered as charged. For the specific interactions, the non-local portions of the off-lattice Go-model were used to model specific DNA-protein interactions, since no DNA and protein atoms are covalently bonded. The potential parameters were obtained from the defaults of the Cafemol methods, and typical biological defaults were used. i.e., temperature, 300 K, and the ionic strength, 0.15 M. The initial configurations of the simulations are the native configuration obtained from the crystal structures discussed above, and each of the 5 simulations performed for each system started from different random velocity distributions. The default timesteps of 20 fs for the protein-DNA complex, and 40 fs for protein-only systems were used, while the use of longer time steps was not investigated. Each system was simulated in five different simulations, each simulation was 3.2 µm in length.

C. Analysis

To detect the flexibility of different regions of the protein, RMSF (root mean square fluctuation of atomic positions) is used. It is defined as: $RMSF = \sqrt{\langle (x_i - \bar{x}_i)^2 \rangle}$, where $\bar{x}_i$ is the average value of $x_i$, and a bracket denotes a time average. In all of the analysis, the simulations are first aligned so as to minimize the RMSD to a common initial frame, so as to remove center-of-mass rotations and translations. Correlated motion is described in terms of correlated fluctuations. The covariance matrix, also known as dynamic cross-correlation matrix (DCCM), is calculated through $DCCM_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}$, where $C_{ij} = \langle (x_i - \bar{x}_i)(x_j - \bar{x}_j) \rangle$. These coefficients measure the linear correlation between two residues. A unitary value indicates a positive correlation (movements in the same direction), while -1 indicates a negative correlation (correlated motions in opposite directions). All the structures are fitted to the initial structure by CA atoms. The free energy surface is calculated via principal component analysis (PCA).\textsuperscript{29} PCA is a well-established method for reducing the dimensionality of a data set. The covariance matrix is constructed based on the Cartesian coordinates of Cα atoms of all conformations with respect to the average structure. The covariance matrix is then diagonalized, whose eigenvectors represent the principal geometrical axes in three-dimensional space and the highest eigenvalues represent the most significant relationship between the dimensions. Here we use the first two principal components, PC1 and PC2, and the free energy is given by $F = -kBT[ln\rho(\text{PC1, PC2})]$, where $\rho$ is the probability distribution of conformations.

III. RESULTS AND DISCUSSION

A. Uninhibited ERG has a large fluctuation in DNA-recognition helix α3

In order to see the dynamical difference among the three constructs, ERGu, ERGi, and ERGi:DNA complex, RMSF calculations were performed. The results are shown in FIG. 1 (a). Overall, ERGu has the highest RMSF, with an average of 1.32 Å, while for ERGi it is 1.25 Å, and ERGi:DNA has the smallest value of 1.16 Å (all calculations exclude the highly flexible end 3 residues). This result is consistent with Regan et. al’s study,\textsuperscript{21} where they found that uninhibited ERG undergoes substantial dynamics on the millisecond-to-microsecond timescale but autoinhibited and DNA-bound ERG do not. The RMSF difference between ERGu and ERGi mainly comes from the residues: TYR354, ASP357 and LYS358, with a difference of 0.12, 0.16 and 0.17 Å. These residues are in the DNA binding domain α3 (shown in FIG. 1 (b)), which interacts with the major groove of DNA. It was shown in Regen et. al’s study that the position of TYR354 is different in bound and unbound ERGi, and Y354F mutant significantly weakens the binding of ERGu to DNA as well as ERGi. From the central structures of the first cluster of the three systems (shown in FIG. 1 (b)), we can not see the position difference of TYR354 among the three conformations. We propose that the weakening of binding to DNA by Y354F mutant is produced by fluctuational difference. ARG337, LYS338 and SER339 also have a larger RMSF in ERGu than that in ERGi, with a difference of 0.16, 0.16 and 0.12 Å. These residues reside at the loop connected to α2. When ERGi binds to
DNA, the DNA binding domain $\alpha_3$ decreases its fluctuation, so as to $\alpha_2$ and $\alpha_4$ that are connected to $\alpha_3$ through loops and $\beta$-sheet, but $\alpha_1$ located near N-terminus increases its fluctuation. It has been found that the DNA-binding domain in general requires more flexibility for high-affinity DNA recognition.\textsuperscript{30} Our results show that besides $\alpha_3$ that exhibits a higher fluctuation in unbound state than in bound state, its nearby regions also have a higher fluctuation in unbound state than in bound state. These suggest that the flexibility in the nearby region also contributes to DNA recognition.

**FIG. 1.** (a) RMSF of the three constructs. (b) The central structures of the first cluster of the three systems are represented as small spheres (a cutoff of 2 Å was used for clustering), which are aligned to the crystal structure of DNA-binding conformation. Yellow is for uninhibited, pink for inhibited while cyan and blue for DNA binding.

**FIG. 2.** Covariance matrix excluding ones greater than -0.1 and less than 0.1. The structures are fitted to the initial structure by CA atoms.
B. Inhibited ERG has a high N-C correlation and DNA-binding change the relative direction of the correlated motion

In order to study the correlated motions, we calculated the covariance matrix which is shown in FIG. 2. The figure presents the main covariance matrix elements by excluding those greater than -0.1 or less than 0.1. As covariances tend towards 1, i.e. completely correlated, the matrix elements become red, while as the covariances tend towards -1, completely anti-correlated, the matrix elements become blue. FIG. 3 shows the differences among the three covariance matrices, which are the covariance differences between their absolute values. This figure only shows the long distance covariance matrix by excluding the covariance between pairs whose distance is smaller than Go potential cutoff of 2.5 Å. Small covariance difference, greater than -0.1 and less than 0.1, are also not shown for clarity. From FIG. 2 and FIG. 3, one finds that ERGi increases anti-correlation of $\alpha_1$-$\alpha_4$, and $\alpha_3$-wing($\beta_3$-$\beta_4$), while DNA-binding decreases correlation of $\alpha_1$-$\alpha_4$, and $\alpha_3$-wing($\beta_3$-$\beta_4$). This result suggests that $\alpha_1$-$\alpha_4$, and $\alpha_3$-wing($\beta_3$-$\beta_4$) correlations disfavor DNA binding. Instead, $\alpha_1$ and $\alpha_3$ have a high correlation with DNA in the complex. Overall, ERGi has the highest anti-correlation of $\alpha_1$-$\alpha_4$ as well as $\alpha_3$-wing($\beta_3$-$\beta_4$). This observation reveals that a general trend of correlation change in the order of: inhibited $>$ uninhibited $>$ DNA-bound. DNA-binding not only decreases $\alpha_1$-$\alpha_4$ correlation regarding its absolute value, but also changes correlation behaviour from negatively-correlated to positively-correlated. Moreover, DNA-binding has a potential to change the $\alpha_3$-$\alpha_4$ covariance from negatively-correlated to positively-correlated. At the same time, it increases the correlation of $\alpha_1$-$\alpha_2$, $\alpha_1$-$\alpha_3$.

Existing studies showed that DNA duplexes bound by ETS domains are typically distorted from an ideal B-DNA geometry, and the interactions between DNA and protein are mostly hydrogen binding between two invariant arginines in the recognition helix H3 and the two guanines.

**FIG. 3.** Covariance matrix difference between (a) inhibited and uninhibited, (b) inhibited and DNA binding, (c) uninhibited and DNA binding. The figure only shows the long distance covariance matrix by excluding the covariance between pairs whose distance is smaller than Go potential cutoff of 2.5 Å. Small covariance difference, greater than -0.1 and less than 0.1, are also not shown for clarity. Here the differences are between the absolute values.
of the GGA(A/T) core. Besides these direct interactions, our correlation matrix (FIG. 2 (d)), however, shows that the residues in the “turn” between helix α2 and α3, the “wing” between strands β3 and β4, and the N terminus of helix H1 have a high correlation with DNA. This is in agreement with existing result that these regions have indirect interactions with DNA.31

C. The three systems share similar conformations

To further look at the structure difference among the three constructs, principle component analysis (PCA) is performed along with a free energy reconstruction. The free energy surfaces projected on the first two eigenvectors from the DNA-binding simulations are shown in Fig. 4. The figure shows that the three constructs share a similar free energy profile, except that the DNA binding simulations sample some rare regions of configuration space in addition to the main well. These results show that the three systems, despite their different dynamics, do have largely similar free energy surfaces, which means that the three systems share similar conformations. This can also be confirmed by the superimposition of the central structure of the first cluster for each system (shown in Fig. 1 (b)). ERGi has a RMSD of 0.39 Å, while ERGu has a RMSD of 0.35 Å with respect to ERGi:DNA.

IV. CONCLUSIONS

By using molecular dynamics simulations, we studied the dynamical properties of ERG protein. After examining the fluctuations and covariance matrices, we find that first, uninhibited ERG has a large fluctuation in DNA-recognition helix α3, which confirms that the DNA-binding domain in general requires more flexibility for high-affinity DNA recognition.30 Besides that, our results show that the high fluctuation of nearby region is also crucial for DNA recognition. Second, Inhibited ERG has a high N-C correlation which contributes to the autoinhibition. Though the N-C coupling contributes to the autoinhibitory in Ets-1, the N terminal inhibitory helices are H1-1 and H1-2,11 which is quite different from ERG, whose N terminal helix is α1 as stated in our work. Once α1 and α4 become uncoupled, ERG tend to bind to DNA, end up of changing the relative direction of correlated motion after DNA binding. Third, the three systems, even though having dramatic different dynamics, yet share similar conformations based on our PCA and structural analysis. Overall, our results support that autoinhibitory of ERG comes from dynamical coupling rather than from conformational change.
ACKNOWLEDGMENTS

All calculations were performed on the DEAC cluster which is supported by WFU Information systems and the WFU Provost’s office. This research was partially supported by NIH RO1CA129373 to FRS.

1. J. Kim and J. Yu, Biochim Biophys Acta ISSN 0006-3002 1825, 186 URL http://www.ncbi.nlm.nih.gov/pubmed/22240201 (2012).
2. E. S. Reddy and V. N. Rao, Oncogene ISSN 0950-9232 6, 2285 (1991) URL http://www.ncbi.nlm.nih.gov/pubmed/1766675.
3. H. R. Siddique, V. N. Rao, L. Lee, and E. S. Reddy, Oncogene ISSN 0950-9232 8, 1751 (1993) URL http://www.ncbi.nlm.nih.gov/pubmed/8510921.
4. K. Murakami, G. Mavrothalassitis, N. K. Bhat, R. J. Fisher, and T. S. Papas, Oncogene ISSN 0950-9232 8, 1559 (1993) URL http://www.ncbi.nlm.nih.gov/pubmed/8502479.
5. S. A. Tomlins, D. R. Rhodes, S. Pernier, S. M. Dhanasekaran, R. Mehra, X.-W. Sun, S. Varambally, X. Cao, J. Tchinda, R. Kueter et al., Science (New York, N.Y.) ISSN 0036-8075 310, 644 (2005).
6. M. a. Pufall and B. J. Graves, Annual review of cell and developmental biology ISSN 1081-0706 18, 421 (2002) URL http://www.ncbi.nlm.nih.gov/pubmed/12142282.
7. B. J. Graves, D. O. Cowley, T. L. Goetz, J. M. Petersen, M. D. Jonsen, and M. E. Gillespie, Cold Spring Harb Symp Quant Biol ISSN 0091-7451 63, 621 (1998) URL http://www.ncbi.nlm.nih.gov/pubmed/10384327.
8. J. Hagman and R. Grosschedl, Proc Natl Acad Sci U S A ISSN 0027-8424 89, 8889 (1992) URL http://www.ncbi.nlm.nih.gov/pubmed/1409581.
9. F. Lim, N. Kraut, J. Frampton, and T. Graf, EMBO J ISSN 0261-4189 11, 643 (1992) URL http://www.ncbi.nlm.nih.gov/pubmed/1311254.
10. M. D. Jonsen, J. M. Petersen, Q. P. Xu, and B. J. Graves, Mol Cell Biol ISSN 0270-7306 16, 2065 (1996) URL http://www.ncbi.nlm.nih.gov/pubmed/8628772.
11. J. J. Skalicky, L. W. Donaldson, J. M. Petersen, B. J. Graves, and L. P. McIntosh, Protein Sci ISSN 0961-8368 5, 296 (1996) URL http://www.ncbi.nlm.nih.gov/pubmed/8745408.
12. J. M. Petersen, J. J. Skalicky, L. W. Donaldson, L. P. McIntosh, T. Alber, and B. J. Graves, Science ISSN 0036-8075 269, 1866 (1995) URL http://www.ncbi.nlm.nih.gov/pubmed/7569926.
13. S. K. Bohlander, Semin Cancer Biol ISSN 1044-579X 15, 162 (2005) URL http://www.ncbi.nlm.nih.gov/pubmed/15828351.
14. H. J. Coyne, S. De, M. Okon, S. M. Green, N. Bhachech, B. J. Graves, and L. P. McIntosh, J Mol Biol ISSN 1098-8638 421, 67 (2012) URL http://www.ncbi.nlm.nih.gov/pubmed/22584210.
15. S. D. A. C. K. Chan, H. J. Coyne, N. Bhachech, U. Hermsdorf, M. Okon, M. E. P. Murphy, B. J. Graves, and L. P. McIntosh, J Mol Biol ISSN 1098-8638 426, 1390 (2014) URL http://www.ncbi.nlm.nih.gov/pubmed/24333486.
16. B. B. Bojović and J. A. Hassell, J Biol Chem ISSN 0021-9258 282, 3027 (2007) URL http://www.ncbi.nlm.nih.gov/pubmed/17148437.
17. A. Greenall, N. Willingham, E. Cheung, D. S. Boam, and A. D. Sharrocks, J Biol Chem ISSN 0021-9258 276, 16207 (2001) URL http://www.ncbi.nlm.nih.gov/pubmed/11278941.
18. A. D. Sharrocks, Nat Rev Mol Cell Biol ISSN 1471-0072 2, 827 (2001) URL http://www.ncbi.nlm.nih.gov/pubmed/11715049.
19. J. Stinson, T. Inoue, P. Yates, A. Clancy, J. D. Norton, and A. D. Sharrocks, Nucleic Acids Res ISSN 1362-4962 31, 4717 (2003) URL http://www.ncbi.nlm.nih.gov/pubmed/12907712.
20. J. L. Kopp, P. J. Wilder, M. Desler, L. Kinarsky, and A. Rizzino, J Biol Chem ISSN 0021-9258 282, 3027 (2007) URL http://www.ncbi.nlm.nih.gov/pubmed/17148437.
21. M. C. Regan, P. S. Horanyi, E. E. Pryor, Jr., J. L. Sarver, D. S. Cafiso, and J. H. Bushweller, Proc Natl Acad Sci U S A ISSN 1091-6490 110, 13374 (2013).
22. T. Terakawa and S. Takada, Biophys J ISSN 1542-0086 101, 1450 (2011) URL http://www.ncbi.nlm.nih.gov/pubmed/21343426.
23. T. Terakawa, H. Kenzaki, and S. Takada, J Am Chem Soc ISSN 1520-5126 134, 14555 (2012) URL http://www.ncbi.nlm.nih.gov/pubmed/22880817.
24. W. Li, T. Terakawa, W. Wang, and S. Takada, Proc Natl Acad Sci U S A ISSN 1091-6490 109, 17789 (2012) URL http://www.ncbi.nlm.nih.gov/pubmed/22753508.
25. K.-i. Okazaki, T. Sato, and M. Takano, J Am Chem Soc ISSN 1520-5126 134, 8918 (2012) URL http://www.ncbi.nlm.nih.gov/pubmed/22559201.
26. R. Kanada, T. Kurosaka, H. Kenzaki, and S. Takada, PLoS Comput Biol ISSN 1553-7358 9, e1002907 (2013) URL http://www.ncbi.nlm.nih.gov/pubmed/23459019.
27. Y. Lu and F. R. Salisbury, Jr., (2014).
28. H. Kenzaki, N. Koga, N. Hori, R. Kanada, W. Li, K.-i. Okazaki, X.-Q. Yao, and S. Takada, Journal of Chemical Theory and Computation ISSN 1549-9618 7, 1979 (2011) URL http://pubs.acs.org/doi/abs/10.1021/ct2001045.
29. A. Amadei, A. B. M. Linsen, and H. J. C. Berendsen, Proteins: Structure, Function and Genetics ISSN 08873585 17, 412 (1993).
30. C. G. Kalodimos, N. Biris, A. M. J. J. Bonvin, M. M. Levandoski, M. Guennuegues, R. Boelens, and R. Kaptein, Science ISSN 1095-9203 305, 386 (2004) URL http://www.ncbi.nlm.nih.gov/pubmed/15256668.
31. P. C. Hollenhorst, L. P. McIntosh, and B. J. Graves, Annu Rev Biochem ISSN 1545-4509 80, 437 (2011) URL http://www.ncbi.nlm.nih.gov/pubmed/21548782.