Published in final edited form as:

Nat Struct Mol Biol. 2009 November ; 16(11): 1118–1120. doi:10.1038/nsmb1109-1118.

How do transcription factors select specific binding sites in the genome?

Yongping Pan, Chung-Jung Tsai, and Buyong Ma

Basic Science Program, SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, National Cancer Institute–Frederick, Frederick, Maryland, USA.

Ruth Nussinov

Basic Science Program, SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, National Cancer Institute–Frederick, Frederick, Maryland, USA.

Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel.

Abstract

How does a transcription factor select a specific DNA response element given the presence of degenerate sequences? To date, this question has largely been viewed from the standpoint of DNA sequence variability and transcription factor binding affinity under steady-state conditions. Here we propose that to address this problem, it is also necessary to account for fluctuating cellular conditions. These lead to dynamic changes in the ensemble of protein (and DNA) conformational states via allosteric effects.

Transcription activation, repression and modulation rest on the binding of transcriptional control proteins to their specific DNA binding sites. Given that the DNA sequences of transcription factor binding sites—the response elements (REs)—are degenerate, how does a transcription factor ‘select’ specific binding sites among the numerous similar binding sites present in the genome? Appropriate RE selection is vital because through this choice the protein can activate, repress or modulate the expression of different genes¹. If the affinity for a specific RE sequence is strong, the transcription factor will bind it; however, the differences in affinities among REs are often minor. Current attempts to address this question have largely focused on RE sequence variability. Conceptually, such an explanation implicitly assumes that RE selection is self-contained and independent of potentially drastic changes in cellular environment. Yet, RE selection is a major means through which cells respond and adapt to changes, and consequently the fluctuating cellular environment should be a key factor in the selection of specific REs. Cells respond to environmental changes by altering the concentrations of specific proteins (and RNAs) and their patterns of post-translational modifications. Because these proteins function by binding to partner proteins such as transcription factors and allosterically changing the conformations of these partners, changes in the relative concentrations of the original proteins will affect the distributions of their partners’ conformations^{2–6}.

Taking all of these points into consideration, we argue here for a ‘different’ view of RE recognition and selectivity: transcription factor proteins exist as dynamic ensembles of

molecules in different conformational states, with different DNA-binding properties. Fluctuating cellular conditions alter the relative concentrations of protein factors that bind to the protein transcription factor. These changes then, through allosteric regulation, affect the relative distribution of the individual conformational states in the ensemble, increasing (or decreasing) certain conformations of the transcription factor^{7,8}; the presence of a higher population of a transcription factor conformation that is compatible with the sequence of certain RE will direct the transcription factor to bind to it²⁻⁶, whereas a higher population of another conformation will direct it to another RE. Similar arguments hold for DNA, which also exists in a dynamic ensemble: fluctuating cellular conditions may lead to a higher concentration of a second transcription factor that binds the DNA near⁹ (or perhaps farther away from) the RE of the first transcription factor. DNA binding by this second transcription factor will allosterically affect the relative distribution of the RE ensemble.

Numerous studies have addressed the relationship between RE sequences, their affinities and their associated functions. These efforts did not, however, provide a clear picture; first, even the gold standard, a consensus DNA sequence, presents a statistical picture; and second, solution binding-affinity experiments are carried out in the absence of other protein factors (Fig. 1a). Cell-based assays can also present a problem: transcription and electrophoretic mobility shift experiments reflect the presence of protein factors at certain steady-state concentrations only (for example, Fig. 1b, or Fig. 1c, or Fig. 1d, but not fluctuating between the figures)—and consequently do not provide information regarding fluctuating cellular conditions. Recent studies of the tumor suppressor p53 are a case in point. *In vitro* experiments measuring the binding affinities of tetrameric p53 for 20 of its response elements (REs) revealed up to 50-fold differences: REs of genes involved in cell cycle arrest and DNA repair and some apoptotic genes bound with higher affinity; and all of the lower-affinity binding sites were in genes involved in apoptosis¹⁰. This implied that although differential affinity is a major factor in functional control, additional factors are also important¹¹. This was supported by observations including the facts that some cells respond to activation by entering cell cycle arrest, whereas others undergo apoptosis; that post-translational modification and subcellular localization of p53 are important in p53-mediated activation and repression; and that some protein factors, such as p300, are critical for certain processes, like cell cycle arrest. Riley *et al.*¹ clearly illustrated a relationship between the REs and p53 function: almost all p53-activated genes present at least one putative DNA-binding site that moderately matches the consensus p53 RE. Yet no clear trend has emerged distinguishing a transcriptional-activator RE sequence from a repressor RE. The estrogen response element (ERE)¹² and the peroxisome proliferator-activated receptors¹³ PPAR- γ and RXR- α provide additional examples. Here we propose that the dynamic change in the concentrations of the protein factors alters the distribution of the transcriptional control protein conformational ensemble, leading to preferential RE binding. We further note that cellular conditions such as stress may also be key to specificity, as with p53, and we emphasize that our focus here is on only one mechanistic transcriptional RE-selection scenario.

We call the different protein factors that bind directly to the transcription control protein (TC) A₁, A₂, A₃, etc., and those interacting indirectly B₁, B₂, B₃, etc. The TC preferentially binds RE1 (Fig. 1a). *In vivo*, there are large fluctuations in A₁, A₂, A₃ and B₁, B₂, B₃.

Binding of A_1 ($A_2, \dots$) to TC allosterically shifts its ensemble (Fig. 2)¹⁴ toward specific RE binding-compatible conformers (in Fig. 1b, preferred binding to RE3; in Fig. 1c, to RE2). Binding reflects protein factor concentration, affinity and TC conformer concentration. This illustrates the problems of extending conclusions from *in vitro* affinity assays to *in vivo* conditions. *In vivo*, protein factors undergo large fluctuations responding to cell cycle, stress, repression and activation (fluctuating among Figs. 1b–d); this is not the case *in vitro* (compare Fig. 1b to Fig. 1c or Fig. 1d). In principle, if there is fluctuation, affinity depends on the correlation between the fluctuations of A_1 and A_2 (A_3 , etc.); if there is no fluctuation, affinity should be predictable on the basis of thermodynamics, that is, from steady-state *in vitro* assays. Experiments mimicking *in vivo* scenarios should reproduce the fluctuations if they follow a time-resolved strategy. If the concentrations of all factors are constant, there should be a direct correlation between the affinity and function. Because concentrations fluctuate, however, the basis of selectivity is elusive. The indirect binding of B_1 , $B_2, \dots$ can also exert longer-range allosteric effects, mediated by $A_1, \dots$ (Fig. 1d), and such allosteric signals can travel nanoscale distances¹⁵.

Proteins are dynamic; however, DNA is not rigid either¹⁶. Like those of proteins, DNA metastable states also consist of ensembles of discrete microstates. Dynamic chromatin remodeling events and binding of additional transcription factors, as in the enhanceosome⁹, to *cis*-regulatory DNA sequences upstream and/or downstream can elicit allosteric effects^{16,17} shifting the RE ensemble, as has been shown for the human BAX promoter¹⁸. The REs of transcriptional control proteins are embedded in dynamic enhancers containing binding sites for additional transcription factors⁹ and are thus expected to display allosteric effects. The well-studied virus-inducible IFN- β enhancer presents a striking example: there are no major protein-protein interaction surfaces between the DNA-binding domains, suggesting cooperative DNA perturbation.

We present two examples that may turn out to provide further experimental support for these ideas. p53-induced apoptosis is enhanced in the presence of ASPP1 or ASPP2, increasing the p53 affinity for proapoptotic REs¹⁹. Despite having high sequence similarity to ASPP, iASPP inhibits p53-dependent induction of proapoptotic genes, suggesting a possible inhibition mechanism whereby iASPP alters p53RE binding. A recent NMR study observed that the p53 interaction interface for the proapoptotic ASPP2 is distinct from that for the antiapoptotic iASPP: ASPP2 binds the p53 DNA-binding domain, whereas iASPP largely interacts with a linker region adjacent to this domain²⁰. Binding of ASPP2 to p53 shifts the conformational ensemble of p53 toward a state that is favorable for binding a proapoptotic RE, whereas binding of iASPP shifts the p53 ensemble toward a conformation favorable for binding an antiapoptotic RE. However, further insights into the mechanisms by which ASPP2 acts on p53 are needed to confirm this.

The second example relates to allostery via REs. Hormones regulate the nuclear hormone receptors through dynamic changes in the receptor ensemble²¹. Large cavities in nuclear receptors such as PPARs and LXRs can bind distinct ligands with minor conformational changes, as demonstrated for PPAR- γ ¹³. Upon hormone binding, the receptor associates with high affinity with specific REs. For the glucocorticoid receptor (GR), binding REs

differing by even a single base pair differentially affects the receptor's conformation and function, showing that REs may act as sequence-specific allosteric ligands of GR²².

Increasing complexity in organisms may lead, from an evolutionary standpoint, to a better-orchestrated cellular network. From a cellular standpoint this leads to higher efficiency and tighter control. Yet from the molecular standpoint it presents a challenging problem—and the more proteins a network hub can potentially bind, the more complex the problem becomes: how are selectivity and specificity introduced? This is particularly challenging when the differences in affinity are marginal, a situation often encountered in regulation. The strategy adopted by evolution may be general: conformational dynamics that is modulated by proteins whose concentrations are under tight control. Binding of proteins to ligands always leads to population shifts^{7,8,23}. Isoforms of transcription control proteins may simulate similar effects (Fig. 2). To explain selective RE binding, we propose that mechanistically, protein factors that bind to a transcriptional control protein change its dynamic energy landscape^{3,4,24}, thus increasing populations favored to bind a specific response element. Experiments reproducing large fluctuations in the concentrations of the protein factors may better capture a relationship between these, the RE sequence and functional consequences. We note that, although for simplicity we have focused here on modulation by protein factors, other events are well known to contribute in similar ways, foremost among them post-translational modifications (such as phosphorylation, acetylation and methylation), small-molecule binding and mutational events. Fluctuations and dynamics are key to the understanding of molecular and cellular scale events^{2–4,14,25,26}; and this holds for all processes in living systems.

ACKNOWLEDGMENTS

This project was funded in whole or in part with federal funds from the US National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government. This research was supported (in part) by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research.

References

1. Riley T, Sontag E, Chen P & Levine A *Nat. Rev. Mol. Cell Biol.* 9, 402–412 (2008). [PubMed: 18431400]
2. Ma B, Kumar S, Tsai CJ & Nussinov R *Protein Eng.* 12, 713–720 (1999). [PubMed: 10506280]
3. Kumar S, Ma B, Tsai CJ, Sinha N & Nussinov R *Protein Sci.* 9, 10–19 (2000). [PubMed: 10739242]
4. Tsai CJ, Ma B & Nussinov R *Proc. Natl. Acad. Sci. USA* 96, 9970–9972 (1999). [PubMed: 10468538]
5. Lange OF et al. *Science* 320, 1471–1475 (2008). [PubMed: 18556554]
6. Boehr DD & Wright PE *Science* 320, 1429–1430 (2008). [PubMed: 18556537]
7. Gunasekaran K, Ma B & Nussinov R *Proteins* 57, 433–443 (2004). [PubMed: 15382234]
8. del Sol A, Tsai CJ, Ma B & Nussinov R *Structure* 17, 1042–1050 (2009). [PubMed: 19679084]
9. Panne D *Curr. Opin. Struct. Biol.* 18, 236–242 (2008). [PubMed: 18206362]
10. Weinberg RL, Veprintsev DB, Bycroft M & Fersht AR *J. Mol. Biol.* 348, 589–596 (2005). [PubMed: 15826656]
11. Espinosa JM *Oncogene* 27, 4013–4023 (2008). [PubMed: 18278067]

12. El Marzouk S, Gahattamaneni R, Joshi SR & Scovell WM J. Steroid Biochem. Mol. Biol 110, 186–195 (2008). [PubMed: 18479910]
13. Chandra V et al. Nature 456, 350–356 (2008). [PubMed: 19043829]
14. Boehr DD, Nussinov R & Wright PE Nat. Chem. Biol 5, 789–796 (2009). [PubMed: 19841628]
15. Ma B & Nussinov R Proc. Natl. Acad. Sci. USA 106, 6887–6888 (2009). [PubMed: 19416924]
16. Völker J, Klump HH & Breslauer KJ Proc. Natl. Acad. Sci. USA 105, 18326–18330 (2008). [PubMed: 19015511]
17. Chaires JB ACS Chem. Biol 3, 207–209 (2008). [PubMed: 18422302]
18. Thornborrow EC & Manfredi JJ J. Biol. Chem 276, 15598–15608 (2001). [PubMed: 11278953]
19. Sullivan A & Lu X Br. J. Cancer 96, 196–200 (2007). [PubMed: 17211478]
20. Ahn J, Byeon IJ, Byeon CH & Gronenborn AM J. Biol. Chem 284, 13812–13822 (2009). [PubMed: 19246451]
21. Nagy L & Schwabe JW Trends Biochem. Sci 29, 317–324 (2004). [PubMed: 15276186]
22. Meijsing SH et al. Science 324, 407–410 (2009). [PubMed: 19372434]
23. Tsai CJ, Del Sol A & Nussinov R Mol. Biosyst 5, 207–216 (2009). [PubMed: 19225609]
24. Frauenfelder H, Sligar SG & Wolynes PG Science 254, 1598–1603 (1991). [PubMed: 1749933]
25. Smock RG & Gierasch LM Science 324, 198–203 (2009). [PubMed: 19359576]
26. Liu J & Nussinov R PLoS Comput. Biol 5, e1000527 (2009). [PubMed: 19798438]

Figure 1.

Schematic illustration of selective binding of transcriptional control proteins to their degenerate response elements under different environmental conditions. TC is the transcriptional control protein; RE1, RE2,... are similar response elements with marginally different binding affinities to TC; A1, A2,... are proteins that interact directly with TC; B1, B2,... are proteins that interact indirectly. **(a)** Test tube scenario: a preferred TC conformation selectively binds RE1. **(b)** High concentration of factor A1; A1 allosterically changes the ensemble of TC conformations, increasing a population favored to bind RE3. **(c)** High concentration of A2; A2 allosterically changes the ensemble of TC conformations, increasing an RE2-favored population. **(d)** High concentrations of A2 and B2; B2 binds to A2, allosterically altering the A2 ensemble, and in turn the TC ensemble, enhancing an RE4-favored state. In cell-based assays A1, A2, B1, B2, etc. are present; however, their concentrations are unchanged, unlike *in vivo*. We focused here on protein allostery, but DNA and chromatin dynamic effects also have crucial roles in the selectivity of the transcriptional factor–DNA binding, emphasizing the importance of the RE location.

Figure 2.

Schematic illustration of the free-energy-landscape shifts associated with the binding of the transcriptional control protein (TC) to various degenerate response elements (RE1, RE2, RE3 and RE4). This diagram illustrates the allosteric effects of the protein factors on TC. Notations for degenerate response elements and proteins factors are as in Figure 1. On the one hand, at top left, the energy landscape favors the binding of TC to RE1; however, after binding of A1, a population shift occurs that favors TC's recognizing RE3. On the other hand, binding of the protein factor A2 to TC shifts the energy landscape to favor TC binding to RE2. A further binding event in which protein factor B1 binds to A2 shifts the energy landscape in favor of TC's recognizing response element RE4. All the energy-landscape shifts depicted here occur through minor conformational changes of TC at the RE binding region (not shown) caused by allosteric binding.