Modulating the DNA affinity of Elk-1 with computationally selected mutations

Sheldon Park[1,3], Eric T. Boder[2], and Jeffery G. Saven[1*]

1. Makineni Theoretical Laboratories, Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, Pennsylvania 19104, USA.
2. Department of Chemical and Biomolecular Engineering, University of Pennsylvania, 220 South 33rd Street, Philadelphia, PA 19104.
3. Current Address: 905 Furnas Hall, Department of Chemical and Biological Engineering, SUNY, Buffalo, NY 14260.

*Corresponding author:
Department of Chemistry
University of Pennsylvania
231 South 34th Street
Philadelphia, PA 19104

Tel: 215-573-6062
Fax: 215-573-2112
Email: saven@sas.upenn.edu

Introduction

Understanding how the specificity and strength of protein-DNA interactions are determined is an important biophysical goal. Transcription factors typically bind their target DNA sequences using a separate DNA binding domain (DBD) within the protein. Despite detailed studies of a diverse array of DNA binding proteins (see [1-3] and references therein), our understanding of sequence specific protein-DNA interactions at the molecular level is far from complete. Protein stability has been shown to be important for DNA binding. For example, mutations in p53 tumor suppressor that destabilize the protein structure also compromise its ability to bind DNA [4]. A recent study comparing unrelated DNA binding proteins has suggested that the strength of protein-DNA interactions is determined by a combination of enthalpy and entropy and these thermodynamic contributions are coupled to ensure sufficient binding affinity [5]. Together, they suggest that studies of both bound and unbound forms of a DBD may be necessary to fully understand the thermodynamics of protein-DNA interaction.

The Ets protein family contains ~50 metazoan transcription factors involved in development, differentiation and proliferation [6]. An Ets protein PEA3, for example, can down regulate the expression of the *HER-2/neu* gene linked to a malignant form of breast and ovarian cancer [7, 8]. While their DNA binding activities are regulated through distinct mechanisms [9, 10], Ets proteins all contain a highly conserved DBD comprising ~85 residues (ETS domain) that uses a winged helix-turn-helix motif to recognize DNA (Fig. 1a). Two highly homologous proteins Elk-1 (Elk) and SAP-1 (SAP) share ~80% sequence identity in the ETS domain and each bind the c-Fos promoter sequence in mammalian cells ($P_{c\text{-}fos}$) cooperatively with the serum response

factor (SRF). They also recognize and bind the E74 promoter sequence (P_{E74}) found in Drosophila with high affinity. The structures of SAP bound to P_{E74} and $P_{c\text{-}fos}$ and of Elk bound to P_{E74} have been solved [11, 12]; these structures are highly similar, having an RMSD of 1.2 Å for the Cα atoms.

Although they share many common amino acid side chains, including identical residues on the recognition helix (H3 in Fig. 1b), in the absence of SRF, Elk and SAP display different DNA sequence specificity [13]. SAP binds both P_{E74} and $P_{c\text{-}fos}$ tightly as a monomer, whereas Elk only binds P_{E74} with high affinity [14]. The high sequence and structural similarity between the two proteins makes the pair an interesting model system for studying the origin and modulation of DNA affinity. Biochemical and structural studies in the past [11-17] have not fully accounted for the observed variations in the binding characteristics of Elk and SAP. The structural properties of the unbound proteins have not been reported and may be responsible for their divergent DNA affinities. In particular, differences in the stability of select key elements in unbound proteins and/or their structural pre-organization prior to DNA binding, which are difficult to observe even in high resolution structures, may significantly alter the strength of their interactions with DNA.

To understand the relationship between protein stability and DNA binding as well as the role of structural organization in DNA binding, we performed 4 ns MD simulations of the ETS domains of Elk and SAP. We also studied the DNA binding activity of a series of rationally constructed Elk mutants. We show that introducing a point mutation in Elk modulates its affinity for $P_{c\text{-}fos}$ to the same degree by which the protein's estimated stability is altered. Our study demonstrates that protein stability and flexibility are integral factors in determining the strength of protein-DNA interaction.

Discussion

The current study uses MD simulations and an activity assay to probe the activity, affinity and sequence specificity of a DNA binding protein. Reducing the internal degrees of freedom has been shown to strengthen macromolecular interactions [18] and has led to developments of high affinity ligands with conformational constraints [19, 20]. On the other hand, the relationship between structural disorder and DNA affinity has not been demonstrated for a natural transcription factor. To examine the role of structural fluctuations in the unbound transcription factor, we simulated the dynamics of two highly homologous transcription factors, Elk and SAP. The sequence and structural similarity between the two proteins allows their divergent DNA binding characteristics to be understood from their individual biophysical properties. The simulation studies show that Elk, with its significantly higher backbone RMSD, is intrinsically more dynamic than SAP. Sequence-specific interactions between proteins and DNA often involve conformational changes in the protein and result in a loss of entropy due to local folding of the binding surface [21]. As such, stabilization of the binding surface is likely an important contributor to high-affinity binding since the entropic cost of association is reduced by pre-organizing the binding surface. Similarly, the 2.2 Å resolution structure of DNA bound p53 shows that some cancer-causing mutations in p53 map to its DNA binding surface, where mutations of R175, R249, R282 and G245 inactivate the protein by destabilizing the binding surface due to lost hydrogen bonds to main chain carbonyl groups [22]. The results of both the

MD simulations and the reporter activity assay are consistent with the interpretation that mutations stabilizing the recognition helix H3 of Elk lead to increased binding affinity for $P_{c\text{-}fos}$.

The change in stability computed using known helix propensities of various amino acids and charge-macrodipole interaction correlates well with measured changes in transcriptional activity. In general the activity of an Elk mutant for $P_{c\text{-}fos}$ is higher when D69 is substituted with a residue with higher helix propensity, which suggests that DNA affinity may be coupled to protein flexibility and stability. This coupling between structure and function provides a plausible model for the role of D69 of Elk in DNA binding [14]. Based on the two observed conformations for D69, Mo et al. concluded that this residue may affect the DNA binding of Elk by reorienting K70, which in turn interacts with Y66 [12]. The corresponding residue K69 in SAP instead forms a salt bridge to the phosphate backbone [11]. However, as the authors also noted, the conformation of D69 that allows a salt-bridge formation is populated less frequently than the alternate conformation that is inconsistent with a salt bridge, and the postulated aliphatic side chain interactions between D69 and K70 in Elk does not explain the absence of similar interactions between V68 and K69 in SAP. The model proposed here does not invoke specific side chain-side chain interactions but instead postulates that residue 69 influences DNA binding by altering the local stability of the backbone. At the same time, this simpler model does not rule out potential interactions involving side chains. Further structural and dynamics studies will be useful for discerning the relative importance of networks of inter-residue interactions in DNA-binding proteins.

The intrinsic flexibility of Elk is concentrated in three separate regions rather than uniformly distributed throughout the protein. Yet stabilizing R1-R3 is unlikely to have the same effect on DNA binding and transcriptional activity, since mutations that stabilize the bound and unbound protein conformations equally do not change the free energy difference and thus make null contributions to the overall binding affinity. The large deviation of simulated structures from the crystal structure in R1 may be due to crystal packing, which is also thought to be responsible for the structural differences between Elk and SAP in the region [12]. Our decision to focus on R3 and mutate D69 for experimental studies derives from its proximity to the DNA recognition helix, where such a mutation could exert a maximum effect on DNA binding. The RMSD disparity in simulated structures of Elk and SAP also coincides with the sequence difference at residue 69, suggesting that flexibility in the region may be a critical element in fine-tuning the DNA affinity. The simulations show that the *i* to *i+3* or *i* to *i+4* backbone hydrogen bonds from V68 and K69 to Y65 of SAP frequently occur throughout the simulation, whereas the corresponding hydrogen bonds in Elk are sparse and intermittent at best. The D69S mutant has roughly 4.7 fold greater affinity for $P_{c\text{-}fos}$ compared to wild type and exhibits a somewhat intermediate hydrogen bonding frequency for Y66. Given that Y66 makes both base and phosphate contacts within the recognition sequence in the Elk-P_{E74} structure, the additional hydrogen bonds to Y65 in SAP and Y66 of D69S thus may facilitate DNA binding by pre-organizing the local structure for DNA association. We also observed during the simulations that the side chain amine of Elk K70 makes occasional hydrogen bonds with the hydroxyl group of Y66. The frequency of this interaction is reduced in the D69S mutant and is absent in SAP. The formation of this fortuitous hydrogen bond with a neighboring side chain may prevent K70 from forming other favorable interactions with DNA and thus further weaken the protein-DNA complex.

The inferred binding free energy gain correlates with predicted stability differences (see Fig. 7). The best linear fit between the increase in protein stability and the increase in binding affinity is obtained with the slope of 1.14. In order to determine the optimum value for the macrodipole moment term, we fitted the subset of mutants not including D69E and parameterized $\Delta\Delta C$ to obtain the y-intercept equal to zero. The destabilization due to the negative charge at D69 was thus estimated to be $\Delta\Delta C = 0.36$ kcal/mol. A closer look at Fig. 7 reveals some deviations from a perfect correlation, most notably for the D69A mutant. One explanation is that the highly exposed position at D69 favors polar side chains such as S and Q over A. Yet D69V does not seem to suffer from the same destabilizing effect (Table 1). Instead, the position-independent helix propensities used in this study may be inaccurate at residue 69. Monte Carlo studies have correlated amino acid helix propensity with side chain entropy loss upon helix formation [23, 24]. Given the proximity to the helix terminus, the entropic advantage of having a short side chain at D69 may be overvalued. A similar observation was made during a mutational study of T4 lysozyme [25], in which the A49S mutation destabilized the protein by 0.5 kcal/mol as expected from reported helix propensities but the A134S mutation destabilized the protein only by 0.1 kcal/mol. The authors attributed the discrepancy to the location of the mutated residues—A49 is located centrally within an α-helix whereas A134 is located towards the carboxy terminus of a helix. For Elk, the lower than expected activity of D69A is consistent with the observation that the intrinsic helix propensities of amino acids may vary depending on their location on the helix [26].

Crystal structures, while rich in information, do not fully address the dynamic nature of protein-DNA interaction. MD simulation studies can offer insight that is difficult to obtain otherwise. Recently, Bruice and coworkers reported MD simulation results of Ets-1 bound to two different DNA sequences containing either GGAA or GGAG at the center, to understand how a single protein can differentiate two related DNA sequences [27, 28]. They postulated that the binding affinity of Ets for the GGAG sequence is lower because of the alternate hydrogen bonds from Y395 (corresponding to Y66 of Elk) to either A3 or C4', which destabilize the bidentate hydrogen bonds from R341 and R344 to DNA bases G2 and G1, respectively. When bound to a high affinity site containing GGAA, the movement of Y395 is restricted by the C5 methyl group of T3', which helps immobilize key hydrogen bonding networks. Analogously, we can learn how two related proteins recognize the same DNA sequence with differing affinity by simulating the dynamics of these proteins and comparing their backbone flexibility. The simulated dynamics of Elk and SAP suggest that excess flexibility within the binding surface of a protein correlates with weaker activity and DNA binding—in agreement with the conclusion from the Ets-1 simulation study that excess mobility of key side chains destabilizes DNA association [27]. The reduced main chain flexibility in SAP therefore appears to be an important contributor to its high affinity, whereas the mobility near the recognition helix seems to disfavor DNA association by Elk. The coupling between structural stability and DNA affinity suggests that, as for many protein-DNA interactions [21], DNA association is an endothermic reaction for Elk involving a local folding of the protein.

Figures

Figure 1

Figure 1a.

The structure of Elk bound to DNA containing the P_{E74} sequence (PDB: 1DUX). The image was generated using SwissPDB viewer and POV-ray v. 3.5.

Figure 1b.

Sequence alignment of the ETS domains of Elk (residues 5-90) and SAP (residues 5-89). Identical residues are indicated with vertical bars, high sequence similarity with two dots, and weak sequence similarity with one dot. The secondary structures of Elk shown above the sequence correspond to helices H1-H3 (bars) and strands S1-S4 (arrows). Residue 69 is in bold. NB. The residue numbering is for Elk.

Figure 1c.

The P_{E74} and $P_{c\text{-}fos}$ sequences used in the study. The two flanking bases that are different are highlighted.

Figure 2

Figure 2a

The main chain RMSD (Å) of Elk and SAP from 4 ns MD simulations.

Figure 2b

The residue-specific RMSD (Å) of Elk (diamond) and SAP (square), time-averaged over the period of 1-4 ns of simulation. The three regions (R1-R3) where the Elk RMSD significantly exceeds that of SAP are indicated. The configurations from two independent simulations were analyzed.

Figure 2c

The RMSD of Elk (diamond) and SAP (square) computed from the B-factors in the Elk-P_{E74} and SAP-P_{E74} structures using $B = \frac{8}{3}\pi^2 RMSD_B{}^2$.

Figure 3

Figure 3a.

The reporter activity of SAP against reporter plasmids containing either nine copies of P_{E74} (filled) or $P_{c\text{-}fos}$ (light) in the promoter. The background (dark) corresponds to yeast transformed with the $P_{c\text{-}fos}$-EGFP reporter alone.

Figure 3b.

The reporter activity of Elk against P_{E74}-EGFP and $P_{c\text{-}fos}$-EGFP. Coloring is the same as in **a**.

Figure 4

The reporter activity of various point mutants: wild type (black), D69V (cyan), D69A (pink), D69S (green), and background (orange).

Figure 5

D69S was analyzed by MD to determine if an increase in DNA affinity correlates with a decreased RMSD. The residue specific RMSD of D69S (green square) was computed by averaging over the 1-4 ns interval of simulation. Ensemble average of two independent simulations.

Figure 6

An example of configurations showing hydrogen bonds formed near the carboxy terminus of the recognition helix. (**a**) SAP, (**b**) Elk, (**c**) D69S.

Figure 7

$\Delta\Delta G_{exp}$ vs. $\Delta\Delta G_{theo}$. The linear fit through the data is $\Delta\Delta G_{exp} = 1.147\ \Delta\Delta G_{theo} - 0.012$ kcal/mol.

Table 1

The normalized fluorescence of single point Elk mutants and the corresponding experimental free energy gain ($\Delta\Delta G_{exp} = 0.6 \ln(F_{mt} / F_{wt})$). The theoretical free energy differences ($\Delta\Delta G_{theo}$) were computed from helix propensities and the macrodipole-side chain interactions (see text).

Figures

Figure 1

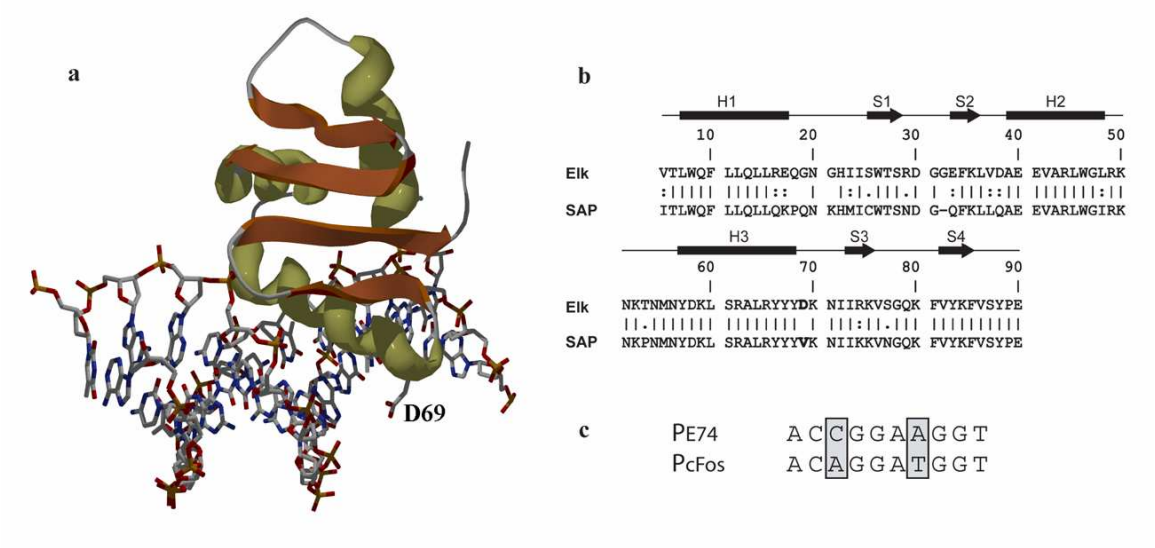

Figure 2

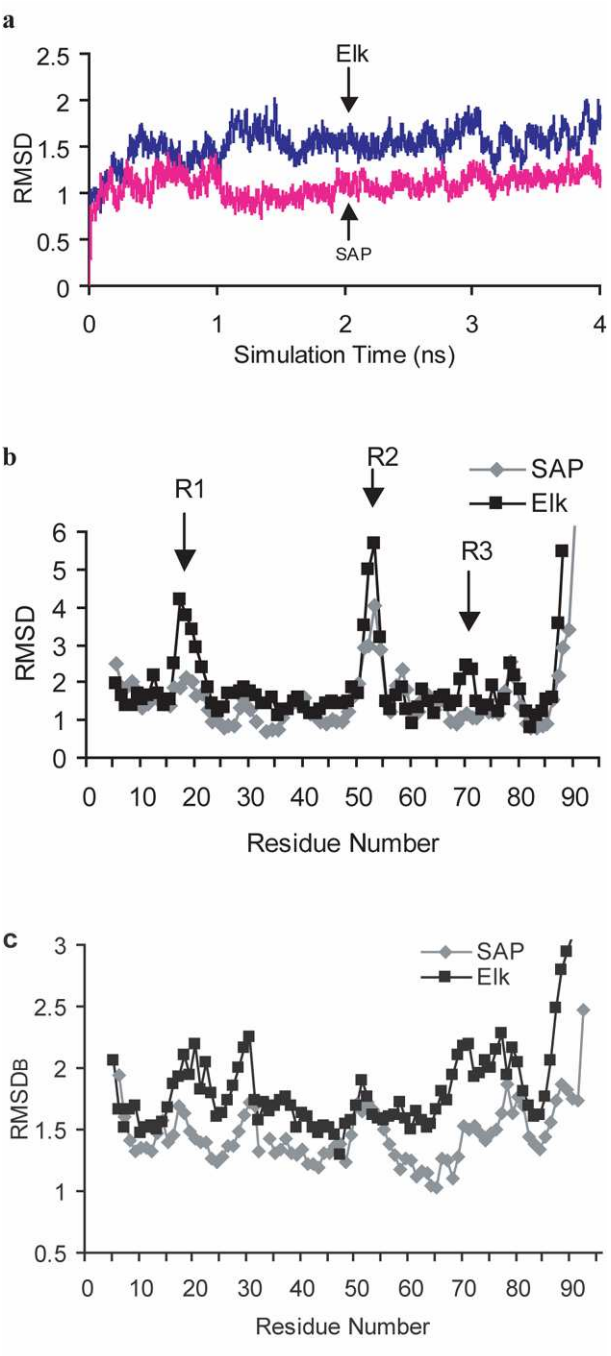
a
Elk
SAP
RMSD
Simulation Time (ns)
b
R1
R2
R3
SAP
Elk
RMSD
Residue Number
c
SAP
Elk
RMSDB
Residue Number

Figure 3

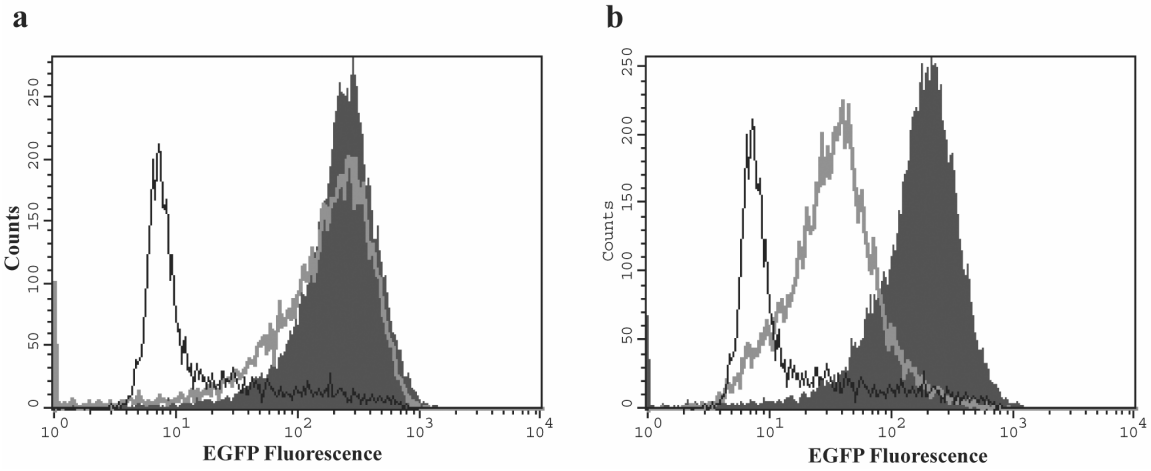
a
b
Counts
EGFP Fluorescence
0
50
100
150
200
250
10^0
10^1
10^2
10^3
10^4

Figure 4

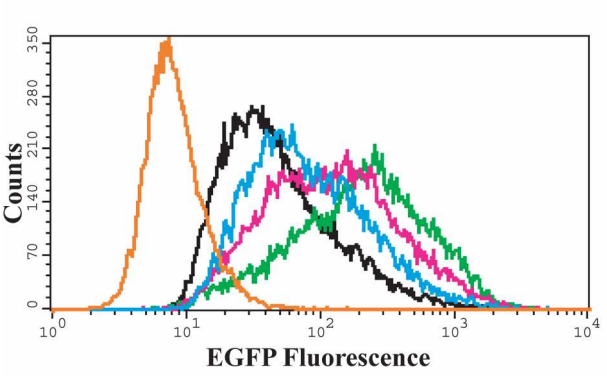
Counts
0
70
140
210
280
350
10^0
10^1
10^2
10^3
10^4
EGFP Fluorescence

Figure 5

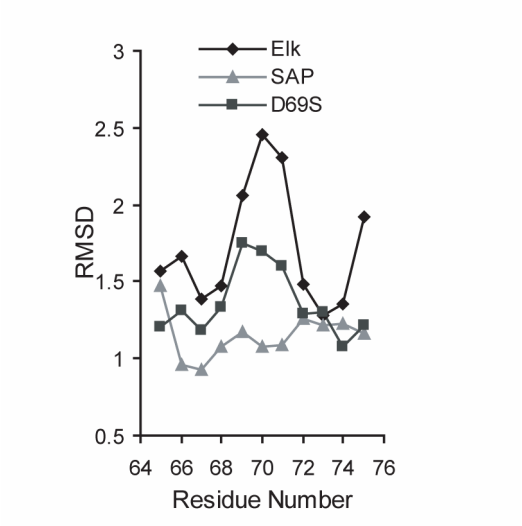
Elk
SAP
D69S
RMSD
3
2.5
2
1.5
1
0.5
64
66
68
70
72
74
76
Residue Number

Figure 6

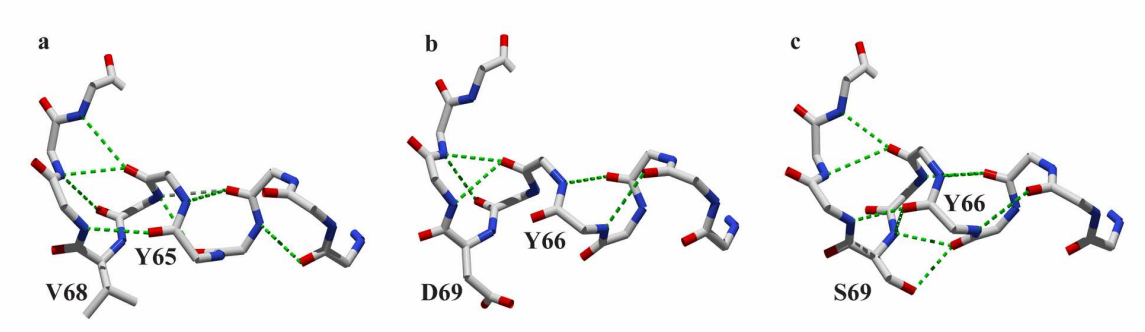

Figure 7

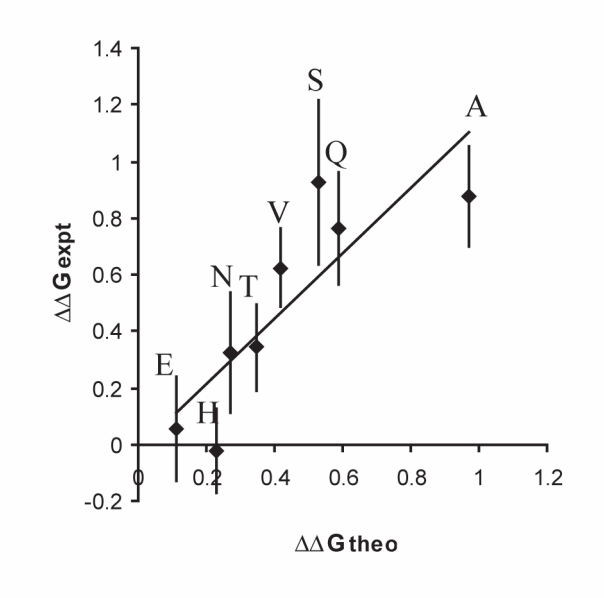
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
ΔΔGexpt
ΔΔGtheo
S
A
Q
V
N
T
E
H

Table 1

	F_{mt}/F_{wt} †	$\Delta\Delta G_{exp}$ (kcal/mol)	$\Delta\Delta G_{theo}$ (kcal/mol)
S	4.7 ± 2.3	0.93 ± 0.29	0.53
A	4.3 ± 1.3	0.88 ± 0.18	0.97
Q	3.6 ± 1.2	0.76 ± 0.20	0.59
V	2.8 ± 0.64	0.62 ± 0.14	0.42
T	1.8 ± 0.45	0.34 ± 0.15	0.35
N	1.7 ± 0.61	0.32 ± 0.21	0.27
E	1.1 ± 0.34	0.055 ± 0.19	0.11
H	0.96 ± 0.24	-0.024 ± 0.15	0.23
SAP	10.1 ± 1.8	1.38 ± 0.11	=

†n=7 for S, A, Q, V, N, E, H. n=4 for T. n=3 for SAP.

References

1. Garvie, C.W. & Wolberger, C. (2001). Recognition of specific DNA sequences. *Mol Cell* **8**, pp. 937-946.
2. Laity, J.H., Lee, B.M. & Wright, P.E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol* **11**, pp. 39-46.
3. Jones, S., van Heyningen, P., Berman, H.M. & Thornton, J.M. (1999). Protein-DNA interactions: A structural analysis. *J Mol Biol* **287**, pp. 877-896.
4. Hainaut, P. & Hollstein, M. (2000). p53 and human cancer: the first ten thousand mutations. *Adv Cancer Res* **77**, pp. 81-137.
5. Jen-Jacobson, L., Engler, L.E. & Jacobson, L.A. (2000). Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure Fold Des* **8**, pp. 1015-1023.
6. Wasylyk, B., Hahn, S.L. & Giovane, A. (1993). The Ets family of transcription factors. *Eur J Biochem* **211**, pp. 7-18.
7. Slamon, D.J., Godolphin, W., Jones, L.A., Holt, J.A., Wong, S.G., Keith, D.E., Levin, W.J., Stuart, S.G., Udove, J., Ullrich, A. & et al. (1989). Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **244**, pp. 707-712.
8. Slamon, D.J., Clark, G.M., Wong, S.G., Levin, W.J., Ullrich, A. & McGuire, W.L. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, pp. 177-182.
9. Sharrocks, A.D. (2001). The ETS-domain transcription factor family. *Nat Rev Mol Cell Biol* **2**, pp. 827-837.
10. Verger, A. & Duterque-Coquillaud, M. (2002). When Ets transcription factors meet their partners. *Bioessays* **24**, pp. 362-370.
11. Mo, Y., Vaessen, B., Johnston, K. & Marmorstein, R. (1998). Structures of SAP-1 bound to DNA targets from the E74 and c-fos promoters: insights into DNA sequence discrimination by Ets proteins. *Mol Cell* **2**, pp. 201-212.
12. Mo, Y., Vaessen, B., Johnston, K. & Marmorstein, R. (2000). Structure of the Elk-1-DNA complex reveals how DNA-distal residues affect ETS domain recognition of DNA. *Nat Struct Biol* **7**, pp. 292-297.
13. Shore, P. & Sharrocks, A.D. (1995). The ETS-domain transcription factors Elk-1 and SAP-1 exhibit differential DNA binding specificities. *Nucleic Acids Res* **23**, pp. 4698-4706.
14. Shore, P., Whitmarsh, A.J., Bhaskaran, R., Davis, R.J., Waltho, J.P. & Sharrocks, A.D. (1996). Determinants of DNA-binding specificity of ETS-domain transcription factors. *Mol Cell Biol* **16**, pp. 3338-3349.
15. Shore, P., Bisset, L., Lakey, J., Waltho, J.P., Virden, R. & Sharrocks, A.D. (1995). Characterization of the Elk-1 ETS DNA-binding domain. *J Biol Chem* **270**, pp. 5805-5811.
16. Bosselut, R., Levin, J., Adjadj, E. & Ghysdael, J. (1993). A single amino-acid substitution in the Ets domain alters core DNA binding specificity of Ets1 to that of the related transcription factors Elf1 and E74. *Nucleic Acids Res* **21**, pp. 5184-5191.
17. Szymczyna, B.R. & Arrowsmith, C.H. (2000). DNA binding specificity studies of four ETS proteins support an indirect read-out mechanism of protein-DNA recognition. *J Biol Chem* **275**, pp. 28363-28370.

18. Khan, A.R., Parrish, J.C., Fraser, M.E., Smith, W.W., Bartlett, P.A. & James, M.N. (1998). Lowering the entropic barrier for binding conformationally flexible inhibitors to enzymes. *Biochemistry* **37**, pp. 16839-16845.
19. Nam, N.H., Ye, G., Sun, G. & Parang, K. (2004). Conformationally constrained peptide analogues of pTyr-Glu-Glu-Ile as inhibitors of the Src SH2 domain binding. *J Med Chem* **47**, pp. 3131-3141.
20. Dinsmore, C.J., Bogusky, M.J., Culberson, J.C., Bergman, J.M., Homnick, C.F., Zartman, C.B., Mosser, S.D., Schaber, M.D., Robinson, R.G., Koblan, K.S., Huber, H.E., Graham, S.L., Hartman, G.D., Huff, J.R. & Williams, T.M. (2001). Conformational restriction of flexible ligands guided by the transferred noe experiment: potent macrocyclic inhibitors of farnesyltransferase. *J Am Chem Soc* **123**, pp. 2107-2108.
21. Spolar, R.S. & Record, M.T., Jr. (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, pp. 777-784.
22. Cho, Y., Gorina, S., Jeffrey, P.D. & Pavletich, N.P. (1994). Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* **265**, pp. 346-355.
23. Creamer, T.P. & Rose, G.D. (1992). Side-chain entropy opposes alpha-helix formation but rationalizes experimentally determined helix-forming propensities. *Proc Natl Acad Sci U S A* **89**, pp. 5937-5941.
24. Creamer, T.P. & Rose, G.D. (1994). Alpha-helix-forming propensities in peptides and proteins. *Proteins* **19**, pp. 85-97.
25. Blaber, M., Lindstrom, J.D., Gassner, N., Xu, J., Heinz, D.W. & Matthews, B.W. (1993). Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala-->Ser and Val-->Thr substitutions in T4 lysozyme. *Biochemistry* **32**, pp. 11363-11373.
26. Engel, D.E. & DeGrado, W.F. (2004). Amino acid propensities are position-dependent throughout the length of alpha-helices. *J Mol Biol* **337**, pp. 1195-1205.
27. Obika, S., Reddy, S.Y. & Bruice, T.C. (2003). Sequence specific DNA binding of Ets-1 transcription factor: molecular dynamics study on the Ets domain--DNA complexes. *J Mol Biol* **331**, pp. 345-359.
28. Reddy, S.Y., Obika, S. & Bruice, T.C. (2003). Conformations and dynamics of Ets-1 ETS domain-DNA complexes. *Proc Natl Acad Sci U S A* **100**, pp. 15475-15480.