Published in final edited form as:

Arch Biochem Biophys. 2014 May 1; 549: 1–11. doi:10.1016/j.abb.2014.03.005.

Role of Promoter DNA Sequence Variations on the Binding of EGR1 Transcription Factor

David C. Mikles, Brett J. Schuchardt, Vikas Bhat, Caleb B. McDonald, and Amjad Farooq*

Department of Biochemistry & Molecular Biology, Leonard Miller School of Medicine, University of Miami, Miami, FL 33136

Abstract

In response to a wide variety of stimuli such as growth factors and hormones, EGR1 transcription factor is rapidly induced and immediately exerts downstream effects central to the maintenance of cellular homeostasis. Herein, our biophysical analysis reveals that DNA sequence variations within the target gene promoters tightly modulate the energetics of binding of EGR1 and that nucleotide substitutions at certain positions are much more detrimental to EGR1-DNA interaction than others. Importantly, the reduction in binding affinity poorly correlates with the loss of enthalpy and gain of entropy—a trend indicative of a complex interplay between underlying thermodynamic factors due to the differential role of water solvent upon nucleotide substitution. We also provide a rationale for the physical basis of the effect of nucleotide substitutions on the EGR1-DNA interaction at atomic level. Taken together, our study bears important implications on understanding the molecular determinants of a key protein-DNA interaction at the cross-roads of human health and disease.

Keywords

Zinc fingers; Protein-DNA thermodynamics; Enthalpy-entropy compensation; Single nucleotide polymorphisms

INTRODUCTION

Binding of transcription factors to DNA in a sequence-specific manner constitutes a key event in regulating signaling networks and henceforth the maintenance of cellular homeostasis. While the classical picture generally portrays the binding of transcription factors to the so-called consensus motifs located within the promoters of target genes, the reality is far from such a simplistic model due to DNA sequence variations. Such changes within the promoter DNA not only add a layer of genetic complexity and diversity but also directly influence its flexibility and its ability to undergo physical phenomena such as bending, stretching, deformation and distortion coupled with its ability to exist in various structural conformations (such as the B-DNA, A-DNA and Z-DNA) [1–3]. Consequently, DNA sequence variations within the promoters play a key role in fine tuning the binding affinity and orientation of transcription factors at the site of DNA. In particular, the

*To whom correspondence should be addressed: amjad@farooqlab.net; 305-243-2429 (tel); 305-243-3955 (fax).

biological activity of a transcription factor at a given promoter and the extent to which it can cross-talk with other cellular factors is highly dependent upon the nature of DNA sequence variations. Given that the outcome of transcriptional machinery is ultimately determined by the level of such cooperation between various transcription factors and co-activators or co-repressors, the role of DNA sequence variations in gauging protein-DNA interactions cannot be overemphasized. Thus, understanding the specificity of protein-DNA interactions must inherently involve integration of the effect of DNA sequence variations on the binding of a transcription factor to a promoter. Toward this goal, we set out here to analyze how DNA sequence variations affect the binding of human EGR1 transcription factor, also known as *Zif268*, to its cognate DNA promoters.

Briefly, EGR1 is comprised of the classical TA-DB modular architecture, where TA is the N-terminal transactivation domain and DB is the C-terminal DNA-binding domain. In response to extracellular stimuli such as hormones, neurotransmitters and growth factors, EGR1 is rapidly induced and exerts its effects at genomic level by virtue of the ability of its DB domain to bind to the promoters of target genes containing the GCGTGGGCG consensus motif, referred to hereinafter as *Zif268* response element (ZRE) (Figure 1), in a sequence-dependent manner. The EGR1-DNA interaction is driven by the binding of DB domain, comprised of three tandem copies of C2H2-type zinc fingers (designated herein ZFI, ZFII and ZFIII), as a monomer to the major groove within the ZRE duplex [4]. Notably, the three zinc fingers act in a cooperative manner to not only impart an arc-like conformation on the DB domain but also enable it to attain a close molecular fit with DNA. The resulting protein-DNA interaction allows the TA domain to recruit various co-activators and cellular factors leading to immediate gene expression responsible for a myriad of cellular activities ranging from cell growth and proliferation to apoptosis and oncogenic transformation [5–8]. Among some of the major targets of EGR1 are genes encoding for tumor suppressors such as PTEN, p53 and p73, growth factors such as TGF β , TNF α and IGFII, and apoptotic regulators Bax and Bcl2 [9–23]. In particular, many of such EGR1-responsive genes harbor not only an impressive array of DNA sequence variations within their ZRE motif but such single nucleotide polymorphisms may also be clinically-relevant [24–30]. A better understanding of the effect of promoter DNA sequence variations on the binding of EGR1 is thus warranted.

In this study, we have conducted a detailed biophysical analysis of the binding of DB domain of EGR1 to all possible single nucleotide variants (SNVs) encompassing the ZRE motif. Our data show that such SNVs tightly modulate the energetics of binding of EGR1 and that nucleotide substitutions at certain positions are much more detrimental to EGR1-DNA interaction than others. The implications of these findings on EGR1-DNA interaction are discussed in molecular terms.

MATERIALS and METHODS

Protein preparation

The DB domain (residues 331–430) of human EGR1 was cloned into pET30 bacterial expression vector with an N-terminal His-tag followed by an enterokinase cleavage site using Novagen LIC technology as described previously [31]. The recombinant protein was

subsequently expressed in *Escherichia coli* BL21*(DE3) bacterial strain and purified on a Ni-NTA affinity column followed by size-exclusion chromatography (SEC) using standard procedures [31]. Final yield was typically between 5–10mg protein of apparent homogeneity per liter of bacterial culture. Protein concentration was spectrophotometrically determined on the basis of an extinction coefficient calculated using the online software ProtParam at ExPasy Server [32]. As reported in our earlier study [31], far-UV CD analysis confirmed that the DB domain possesses an $\alpha\beta$ -fold characteristic of a native fully-folded conformation. It should be noted that the treatment of recombinant DB domain with enterokinase substantially reduced the yield of protein due to partial digestion. While control experiments were carried out on the cleaved construct to check that the His-tag did not alter the properties of the DB domain, the signal-to-noise ratio obtained for these measurements was relatively poor due to the low yield of cleaved protein. Accordingly, all experiments reported herein were conducted on the recombinant DB domain containing an N-terminal His-tag.

DNA synthesis

15-mer DNA oligos containing the ZRE consensus site (GCGTGGGCG) and all possible single nucleotide variants thereof were commercially obtained from Sigma Genosys. The design of such oligos and the nomenclature employed in this study is illustrated in Figure 1. Oligo concentrations were determined spectrophotometrically on the basis of their extinction co-efficients derived from their nucleotide sequence using the online software OligoAnalyzer 3.1 based on the nearest-neighbor model [33]. Double-stranded DNA (dsDNA) oligos were generated as described earlier [31].

ITC measurements

Isothermal titration calorimetry (ITC) experiments were performed on a TA Nano-ITC instrument. Briefly, the DB domain of EGR1 and the dsDNA oligos were dialyzed in 50mM Sodium phosphate, 100mM NaCl and 5mM β -mercaptoethanol at pH 7.0. All experiments were initiated by injecting $25 \times 10\mu\text{l}$ aliquots of 100–200 μM of each dsDNA oligo from the syringe into the calorimetric cell containing 0.95ml of 10–20 μM of DB domain solution at 25°C. The thermal power as a function of each injection was automatically recorded using the integrated NanoAnalyze software. The raw data were further integrated to yield binding isotherms of heat release per injection as a function of molar ratio of each dsDNA oligo to DB domain. The heats of mixing and dilution were subtracted from the heat of binding per injection by carrying out a control experiment in which the same buffer in the calorimetric cell was titrated against each dsDNA oligo in an identical manner. To determine the stoichiometry (n), equilibrium dissociation constant (K_d) and the enthalpy change (ΔH) associated with the binding of DB domain to each dsDNA oligo, the binding isotherms were iteratively fit to a built-in one-site model by non-linear least squares regression analysis using the integrated NanoAnalyze software as described previously [34, 31]. The free energy change (ΔG) upon binding was calculated from the relationship:

$$\Delta G = RT \ln K_d \quad [1]$$

where R is the universal molar gas constant (1.99 cal/mol/K) and T is the absolute temperature (298 K). The entropic contribution ($T \Delta S$) to the free energy of binding was calculated from the relationship:

$$T\Delta S = \Delta H - \Delta G \quad [2]$$

where ΔH and ΔG are as defined above. Heat capacity change (ΔC_p) and enthalpy change at 60°C (ΔH_{60}) associated with the binding of DB domain to ZRE motif and all its variants thereof was determined from the slope of corresponding ΔH - T plot, where T is the temperature.

Molecular modeling

Structural model of the DB domain of EGR1 in complex with 15-mer dsDNA oligo containing the ZRE consensus motif was built using the MODELLER software [35]. Briefly, the crystal structure of DB domain of EGR1 in complex with a dsDNA oligo containing the ZRE consensus motif (PDBID 1ZAA), but with varying flanking sequences, was used as template. A total of 100 structural models were calculated and the structure with the lowest energy, as judged by the MODELLER Objective Function, was selected for further analysis. The structural model was rendered using RIBBONS [36].

SASA calculations

Changes in solvent-accessible surface area (SASA) upon the binding of DB domain of EGR1 to ZRE motif and all its variants thereof were calculated from the corresponding experimentally-determined values of ΔC_p and ΔH_{60} . To determine changes in polar SASA ($\Delta \text{SASA}_{\text{polar}}$) and apolar SASA ($\Delta \text{SASA}_{\text{apolar}}$) upon binding, it was assumed that ΔC_p and ΔH_{60} are additive and linearly depend on the change in $\Delta \text{SASA}_{\text{polar}}$ and $\Delta \text{SASA}_{\text{apolar}}$ as embodied in the following empirically-derived expressions [37–42]:

$$\Delta C_p = a[\Delta \text{SASA}_{\text{polar}}] + b[\Delta \text{SASA}_{\text{apolar}}] \quad [3]$$

$$\Delta H_{60} = c[\Delta \text{SASA}_{\text{polar}}] + d[\Delta \text{SASA}_{\text{apolar}}] \quad [4]$$

where a , b , c and d are empirically-determined coefficients with values of -0.26 cal/mol/K/Å², $+0.45$ cal/mol/K/Å², $+31.34$ cal/mol/Å² and -8.44 cal/mol/Å², respectively. The coefficients a and b are independent of temperature, while c and d refer to a temperature of 60°C, which equates to the median melting temperature of the proteins from which these constants are derived [37, 40, 42]. With ΔC_p and ΔH_{60} experimentally determined using ITC and the knowledge of coefficients a – d from empirical models [37–42], Eqs [3] and [4] were simultaneously solved to obtain the values of $\Delta \text{SASA}_{\text{polar}}$ and $\Delta \text{SASA}_{\text{apolar}}$. Notably, $\Delta \text{SASA}_{\text{polar}}$ and $\Delta \text{SASA}_{\text{apolar}}$ upon binding were also determined from the structural model of the DB domain of EGR1 in complex with the 15-mer ZRE duplex using the following equations:

$$\Delta \text{SASA}_{\text{polar}} = \text{SASA}_{\text{bp}} - (\text{SASA}_{\text{fp}} + \text{SASA}_{\text{dp}}) \quad [5]$$

$$\Delta\text{SASA}_{\text{apolar}} = \text{SASA}_{\text{ba}} - (\text{SASA}_{\text{fa}} + \text{SASA}_{\text{da}}) \quad [6]$$

where SASA_{bp} and SASA_{ba} are the polar and apolar SASA of DB domain bound to DNA, SASA_{fp} and SASA_{fa} are the polar and apolar SASA of DB domain alone, and SASA_{dp} and SASA_{da} are the polar and apolar SASA of ZRE duplex alone. All SASA calculations on the structural model of DB domain bound to DNA were performed using GETAREA with a probe radius of 1.4Å [43]. In all cases, the total change in SASA ($\text{SASA}_{\text{total}}$) is defined by the following equation:

$$\Delta\text{SASA}_{\text{total}} = \Delta\text{SASA}_{\text{polar}} + \Delta\text{SASA}_{\text{apolar}} \quad [7]$$

RESULTS and DISCUSSION

EGR1 binds to ZRE motif and its variants thereof with a broad spectrum of affinities

To understand how nucleotide substitutions affect EGR1-DNA interaction, we analyzed the binding of DB domain of EGR1 to ZRE consensus motif and its single nucleotide variants thereof using ITC (Figure 2 and Table 1). Our analysis reveals that DNA sequence variations within the ZRE motif tightly modulate the energetics of binding of EGR1 and that a handful of single nucleotide substitutions reduce binding affinity by more than an order of magnitude. This observation is consistent with previous studies showing that the promoter DNA sequence variations can dramatically affect EGR1-DNA interaction [44–47]. Importantly, such reduction in binding affinity poorly correlates with the loss of enthalpy and gain of entropy—a trend indicative of a complex interplay between underlying thermodynamic factors as reflected in Figure 3. Thus, for example, the reduction in the binding affinity of T-2 and T+4 variants predominantly results from the loss of favorable enthalpy while overall entropic contribution becomes favorable relative to the binding of ZRE motif in both cases. This is intuitive in the sense that these nucleotide substitutions presumably result in the loss of specific protein-DNA contacts (loss of enthalpy) but, in doing so, allow the protein-DNA complex to exercise greater motional freedom (gain of entropy). On the other hand, enthalpic contribution to the reduced binding of G-1 and T+1 variants is greater than that observed for the binding of ZRE motif but this is largely offset by a much greater entropic contribution. That this is so strongly invokes the role of water solvent in modulating protein-DNA thermodynamics upon the introduction of these nucleotide substitutions. Thus, for example, it is conceivable that the loss of enthalpy due to the loss of protein-DNA contacts upon the introduction of G-1 and T+1 substitutions is more than offset by favorable enthalpic gain resulting from the formation of compensatory protein-water and DNA-water contacts. However, the immobilization of such bulk waters will also likely lead to the loss of additional solvent entropy. Such a scenario could thus account for the thermodynamic profiles observed for the binding of G-1 and T+1 variants to DB domain relative to ZRE motif.

Of particular note is the observation that while the A+2 variant binds to the DB domain with an affinity that is close to 20-fold weaker compared to ZRE motif, their underlying thermodynamic signatures are almost indistinguishable (Table 1 and Figure 3). Contrasting

this observation are the thermodynamic behaviors of T-2 and A+1 variants. Thus, while both of these variants bind to the DB domain with comparable affinities, they do so with remarkably distinct thermodynamic signatures—whereas the binding of T-2 variant is accompanied by both favorable enthalpic and entropic changes, binding of A+1 variant is concomitant with a substantial entropic penalty. In particular, the binding mode of A+1 variant appears to be the more general mechanism as it is also shared by the ZRE consensus motif and a vast majority of other variants. In contrast, T-2 along with T+4 represent the only variant motifs for which both enthalpic and entropic changes are favorable. It should be noted that the favorable enthalpic changes observed here are consistent with the formation of an extensive network of intermolecular hydrogen bonding, ion pairing and van der Waals contacts at the protein-DNA interface [4]. On the other hand, the unfavorable entropic changes most likely result from the loss of conformational degrees of freedom that both the protein and DNA experience upon complexation. Given that the DB domain must adopt an arc-like conformation in complex with DNA [4], the entropic penalty observed here may also in part be attributed to such physical distortion of protein necessary for it to wrap around the DNA so as to attain a close molecular fit.

In sum, our data presented above strongly implicate that DNA sequence variations within the ZRE motif modulate the energetics of binding of DB domain of EGR1 to DNA. Accordingly, DNA sequence variations within the target gene promoters likely play a key role in gauging the transcriptional output of EGR1 in response to a variety of stimuli such as growth factors and hormones. Such DNA sequence variations may have thus evolved to provide a differential response to the expression of EGR1-responsive genes. Importantly, while nearly all of the DNA sequence variations systemically analyzed above are also found within the promoters of a diverse array of EGR1-responsive genes (Table 1), none of these has hitherto been linked to any physiological condition or pathological disorder [24–30].

Physical basis of how DNA sequence variations may alter EGR1-DNA interaction

In an attempt to understand the physical basis of how nucleotide substitutions mitigate EGR1-DNA interaction, we next modeled and analyzed the structure of DB domain in complex with the ZRE consensus motif. As shown in Figure 4, the DB domain of EGR1 is comprised of three tandem copies of C2H2-type zinc fingers, designated herein ZFI, ZFII and ZFIII, which come together in space to assemble into an arc-like conformation that snugly fits into the major groove of DNA. Importantly, the EGR1-DNA interaction is driven by the binding of each zinc finger to one of the three subsites, each subsite being comprised of a trinucleotide sequence, within the 9-bp GCGTGGGCG consensus motif (Figure 1a). Moreover, the EGR1-DNA interaction is stabilized by numerous van der Waals contacts in addition to an extensive network of intermolecular hydrogen bonding and ion pairing. In particular, amino acid residues such as R357, D381, H382 and T385 located within the DB domain play a prominent role in orchestrating such a network of intermolecular contacts (Figure 4). For example, the stacking of imidazole ring of H382 against the pyrimidine ring of T-1 (sense strand) promotes van der Waals contacts between the protein and DNA. This interaction is further buttressed by the van der Waals cooperation between C γ 2-methyl group of T385 and C7-methyl moiety of T-1. On the other hand, O δ 2-hydroxyl oxygen of D381 and N4-ring nitrogen of C-1 (antisense strand) engage in intermolecular hydrogen

bonding. This scenario is also replicated between N η 1-guanidine nitrogen of R357 and N7-ring nitrogen of G+1 (sense strand). How are these key protein-DNA contacts affected upon the introduction of DNA sequence variations, such as those introduced in the G-1 and T+1 variants, within the ZRE motif? It should be recalled that the G-1 and T+1 variants reduce the binding affinity of DB domain to DNA moderately (3-fold) and substantially (25-fold), respectively.

Our structural model suggests that the introduction of a rather bulkier guanine base at the -1 position (G-1) would disrupt van der Waals stacking established between H382 and T-1 within the ZRE motif (Figure 4). Additionally, being devoid of a C γ 2-methyl group, G-1 would be unable to participate in van der Waals contact with T385. The loss of such key intermolecular contacts would thus be expected to result in the loss of favorable enthalpic contribution to free energy and thereby compromise the EGR1-DNA interaction as observed here. Equally importantly, the G-1 substitution would be expected to confer higher rigidity upon DNA by virtue of the ability of G-C base pair to engage in three hydrogen bonds in lieu of two afforded by the T-A base pair in the ZRE motif at the -1 position. Consequently, the loss of DNA flexibility should be expected to not only compromise its ability to attain a closer molecular fit with the protein but also compensate for the loss of enthalpically-favorable protein-DNA contacts through entropic gain. Yet, the thermodynamics associated with the binding of G-1 variant to DB domain paint a totally opposing picture against such a priori considerations (Table 1). Thus, the binding of G-1 variant to DB domain is concomitant with favorable enthalpic gain and unfavorable entropic loss of close to -10 kcal/mol relative to ZRE motif. How can one account for such discrepancy between theory and experiment? As noted above, the binding of ZRE motif and G-1 variant to DB domain likely results in differential solvation of the corresponding protein-DNA complexes. Thus, the water solvent plays a differential role in mediating the binding of ZRE motif and G-1 variant and, in doing so, modulates their binding thermodynamics in a manner that cannot be easily rationalized in terms of the effect of sequence variation on protein-DNA contacts. As discussed later, this is further supported by the observation that the heat capacity change (C_p) associated with the binding of G-1 variant to DB domain is nearly -60 cal/mol/K greater than that observed for the ZRE motif (Table 2).

On the other hand, the effect of replacement of a rather bulkier guanine (G+1) within the ZRE motif with a much smaller and more hydrophobic thymine at the +1 position (T+1) would be expected to be even more dramatic. Indeed, our structural analysis suggests that unlike the ability of G+1 (sense strand) and C-1 (antisense strand) within the ZRE motif to respectively participate in key intermolecular hydrogen bonding contacts with R357 and D381, T+1 within the sense strand and its A-1 counterpart within the antisense strand would fail to do so (Figure 4). In particular, the hydrophobic C γ 2-methyl group of T+1 would be highly destabilizing for subsequent protein-DNA contacts due to its close proximity to the charged guanidine moiety of R357. The rather small size of T+1 compared to a much bulkier G+1 may also result in the formation of cavities and subsequent entrapment of water molecules at the protein-DNA interface. Additionally, the exchange of a G-C base pair with T-A at the +1 position would result in the loss of an intramolecular hydrogen bonding contact and this could also jeopardize the optimal conformation of DNA required to bind to the protein. Given that the binding of T+1 variant to DB domain results in the reduction of

affinity by about 25-fold relative to ZRE motif (Table 1), one would expect the accompanying thermodynamics of binding of T+1 variant to be substantially distinct from those of ZRE motif. Indeed, the binding of T+1 variant to DB domain is enthalpically more favorable accompanied by greater entropic penalty compared to ZRE motif in a manner akin to that observed for the G-1 variant, albeit with a much smaller effect on the thermodynamic parameters (Tables 1 and 2). This further epitomizes the differential role that the water solvent plays in fine-tuning the thermodynamics of macromolecular interactions.

Taken together, our structural analysis discussed above suggests that the loss of key van der Waals and hydrogen bonding contacts would likely mitigate the binding of DB domain of EGR1 to G-1 and T+1 variants in agreement with our thermodynamic data (Table 1). More importantly, these variants exert their effect by virtue of their ability to differentially modulate the solvation of protein-DNA complex relative to ZRE motif. While an exhaustive analysis is beyond the scope of this work, the arguments presented above clearly suggest that the reduction of binding affinity due to other nucleotide substitutions likely results from the loss of specific protein-DNA contacts. However, such loss of intermolecular interactions does not necessarily correlate with the loss of enthalpy and gain of entropy but rather the effect of nucleotide substitutions on the underlying thermodynamics can be best accounted for through the differential contribution of water solvent. It is also important to note that the nucleotide substitutions have little or negligible effect on the B-conformation of ZRE motif as probed by far-UV CD analysis (data not shown) and thus conformational changes within DNA due to sequence variations are unlikely to dramatically account for the reduction in binding affinity observed here.

Binding of EGR1 to ZRE motif and its variants thereof is enthalpy-entropy compensated

Enthalpy-entropy compensation is a thermodynamic bottleneck in that it usually eliminates the effect of enthalpic or entropic advantages on macromolecular interactions. In this ubiquitous thermodynamic phenomenon [48–52], the gain of favorable enthalpy, for example due to the introduction of a new hydrogen bond donor or acceptor, is largely compensated by unfavorable entropic factors and vice versa such that there is little or no change in the overall free energy of binding. While this may appear to be a negative aspect of engineering high-affinity therapeutic inhibitors, on the contrary, it offers biological systems with natural resilience against structural and/or functional distortions due to environmental stresses. Although the molecular origin of enthalpy-entropy compensation is poorly understood, a number of underlying physical factors have been suggested [53]. These include: (i) solvent reorganization; (ii) conformational restriction of atoms within a given system; (iii) limited temperature range over which experiments are feasible thereby limiting the free energy window; (iv) finite heat capacity of system under study; and (v) multiplicity and cooperativity of intermolecular interactions that directly contribute to the overall free energy as well as those that merely govern binding specificity but with little or no contribution to the free energy. Simply put, enthalpy-entropy compensation arises due to the fact that the introduction of a new bonding interaction (gain of enthalpy) is intrinsically accompanied by an increase in the order of the system (loss of entropy) so as to maintain thermodynamic homeostasis.

In order to test the extent to which the binding of ZRE motif and its variants thereof to the DB domain of EGR1 is also subject to enthalpy-entropy compensation, we generated the enthalpy-entropy plot (Figure 5a). It is notable that the binding of ZRE motif and variants to the DB domain is seemingly enthalpy-entropy compensated in agreement with the observation that these underlying thermodynamic factors play an intricate role in dictating the affinity of EGR1-DNA interaction upon the introduction of DNA sequence variations (Table 1 and Figure 3). Consistent with these observations, we also note that the enthalpic (ΔH) and entropic ($T \Delta S$) contributions for the binding of ZRE motif and variants to the DB domain show poor correlation with the overall free energy (ΔG) (Figures 5b and 5c). For example, an increase in favorable ΔH or a decrease in unfavorable $T \Delta S$ does not necessarily lead to an increase in ΔG and vice versa. Given that the binding of proteins to major grooves within the DNA is under enthalpic control [54–60], the negative contribution of entropic penalty to the free energy thus appears to be an equally important regulator of such protein-DNA interactions. Collectively, these findings bear important consequences on the rational design of synthetic oligos and/or small molecule inhibitors designed to compete with gene promoters for the binding of EGR1. Thus, for example, efforts to optimize such inhibitors on the basis of their binding affinity and clinical efficacy may prove to be more rewarding and cost-effective than attempts directed at optimizing their underlying enthalpic and entropic contributions.

Differential binding of EGR1 to DNA is highly dependent upon the position of nucleotide substitution

Next, we analyzed how the position at which a nucleotide is substituted within the ZRE motif dictates the binding of DB domain of EGR1 in thermodynamic terms (Figure 6). It is important to note here that the EGR1-DNA interaction is driven by the binding of each of its three zinc fingers (ZFI-ZFIII) to one of the three subsites, each subsite being comprised of a trinucleotide sequence, within the 9-bp GCGTGGGCG consensus motif (Figures 1 and 4). Remarkably, our analysis suggests that certain positions within this motif appear to be more susceptible to the effect of DNA sequence variations, while others are much more tolerant. Thus, for example, the greatest loss in the free energy (ΔG) of binding consistently correlates with the introduction of DNA sequence variations at the -2 , $+1$ and $+2$ positions (within the sense strand) and it is more or less mirrored by corresponding changes in underlying enthalpic (ΔH) and entropic ($T \Delta S$) factors.

On the other hand, the effect of DNA sequence variations at the -4 , -3 , -1 , 0 , $+3$ and $+4$ positions (within the sense strand) is relatively less detrimental to EGR1-DNA interaction. Notably, the -2 , $+1$ and $+2$ positions are all occupied by a guanine within each of the three subsites that accommodates one of the three zinc fingers of DB domain. However, they do not represent structurally-equivalent positions within each respective subsite. Thus, while -2 and $+1$ positions correspond to the 3'-terminal nucleotide within subsites that would accommodate ZFI and ZFII, the $+2$ position is the 5'-terminal nucleotide within the subsite accommodating ZFIII. This implies that the effect of various DNA sequence variations within the ZRE motif on each of the three zinc fingers within the DB domain is non-symmetrical. Most importantly, these findings are consistent with the crystal structure of the DB domain of EGR1 in complex with ZRE motif [4], wherein guanine bases at the -2 , $+1$

and +2 positions engage in closer intermolecular contacts with the protein in comparison with those at other positions. Taken together, our analysis suggests that the guanine nucleotides at the -2, +1 and +2 positions within the ZRE consensus motif are most critical for the binding of EGR1 and that DNA sequence variations at these positions are likely to be most detrimental.

Effect of solvation on the differential binding of EGR1 to ZRE motif and its variants thereof

In an effort to investigate how nucleotide substitutions affect the change in heat capacity (C_p)—a key thermodynamic parameter that probes the effect of solvation upon protein-DNA complexation—we measured the dependence of enthalpic change (ΔH) associated with the binding of DB domain of EGR1 to ZRE motif and its variants thereof on temperature (T). Importantly, the temperature-dependence of ΔH is related to C_p by Kirchhoff's relationship $C_p = d(\Delta H)/dT$. Accordingly, C_p accompanying the binding of DB domain to ZRE motif and its variants thereof was determined from the corresponding slopes of ΔH -T plots (Table 2). It is telling that a negative value of C_p in the hundreds of cal/mol/K was observed across the board, implying that the binding of DB domain to ZRE motif and its variants thereof predominantly results in the burial of apolar groups over polar surfaces. However, C_p associated with the binding of ZRE motif ranks somewhere in the middle of the spectrum of C_p values observed for other variants, implying that the extent of the burial of apolar surface area poorly correlates with binding affinity. Notably, negative changes in C_p are essentially a consequence of desolvation of protein and DNA surfaces upon binding as well as due to entrapment of waters within cavities at interacting surfaces [61–63]. The fact that C_p widely varies for different variants is evidence that changes within the promoter DNA sequence result in differential changes in the extent to which water is excluded from interacting surfaces within both the protein and DNA upon complexation.

In order to quantify and compare the relative extent of such desolvation of macromolecular surfaces upon the binding of DB domain to ZRE motif and its variants thereof, we next calculated the corresponding changes in solvent-accessible surface area (SASA) using empirically-derived Eqs [3] and [4]. It is noteworthy that these equations are based on protein stability measurements and therefore only strictly applied to the analysis of protein-protein interactions [37–42]. Thus, the applicability of Eqs [3] and [4] to this study is questionable. In particular, while such analysis on protein-DNA interactions has not been hitherto conducted, studies on protein-carbohydrate interactions generate a very different set of parameterization compared to those observed for protein-protein interactions [64]. Nonetheless, we have previously applied the semi-empirical analysis based on Eqs [3] and [4] to protein-DNA interactions with some success [65, 66]. Notwithstanding the above-mentioned limitations, the binding of DB domain to ZRE motif and its variants thereof is accompanied by the burial of polar ($-\Delta \text{SASA}_{\text{polar}}$) and apolar ($-\Delta \text{SASA}_{\text{apolar}}$) SASA of between approximately 800–1900 Å² and 1200–2100 Å², respectively (Table 2). However, $-\Delta \text{SASA}_{\text{apolar}}$ is between 20–30% greater than $-\Delta \text{SASA}_{\text{polar}}$ in the case of ZRE motif and all of its variants thereof in agreement with the notion that the amount of apolar surface buried is always greater than polar groups. Assuming that the protein experiences little or no structural change upon binding to DNA, $-\Delta \text{SASA}_{\text{polar}}$ and $-\Delta \text{SASA}_{\text{apolar}}$ values calculated

from our structural model of DB domain bound to DNA are in the neighborhood of $1500\text{\AA}^2$ and $1000\text{\AA}^2$, respectively. The fact that the protein is experimentally observed to bury a greater amount of apolar surface over polar groups while the opposite trend is observed in silico strongly suggests that the binding of DNA triggers a substantial conformational change within the DB domain.

Given that the DB domain retains a fully-folded structure in the absence of DNA [31], the most likely scenario to account for such conformational change lies in the ability of three zinc fingers (ZFI, ZFII and ZFIII) to undergo quaternary structural change upon binding to DNA. Supporting this argument is the fact that the zinc fingers impart upon the DB domain an arc-like conformation optimally suited to attain a close molecular fit with DNA (Figure 4). Importantly, the solvent-exposed surface of the unliganded DB domain is bolstered by electrostatic polarization, with the inner face of the arc harboring an overall positive charge, while the outer rim is largely neutral. Thus, while the positively-charged inner face would be required for the ability of DB domain to establish a stable interaction with the negatively charged DNA, the rather highly constrained arc-like architecture will likely undergo some sort of structural rearrangement in the absence of DNA so as to minimize electrostatic repulsions between the three tandem zinc fingers. Accordingly, the DB domain must experience a conformational switching upon binding to DNA in agreement with our SASA calculations based on thermodynamic measurements.

Given that cooperativity in proteins is usually associated with structural rearrangement, we believe that the conformational change observed here may also account for the ability of tandem zinc fingers to drive the binding of DB domain to DNA in a cooperative manner. More importantly, cooperativity in proteins underlies their ability to transfer information from one site to another. In this regard, sequence variations introduced within the ZRE motif would be expected to alter the cooperative binding of variant motifs to DB domain. Indeed, our data presented above show that the extent of burial of apolar and polar surfaces is highly sensitive to the DNA sequence variation (Table 2). This implies that the changes observed in the burial of surface area upon the introduction of DNA sequence variations would variably result in the extent to which the DB domain undergoes conformational change upon binding, thereby directly influencing the binding cooperativity.

CONCLUSIONS

The ability of EGR1 transcription factor to mediate a wide array of cellular activities ranging from cell growth and proliferation to apoptosis and oncogenic transformation is well-documented [5–8]. Notably, cellular expression of EGR1 is down-regulated in glioblastoma, lymphoma, and cancers of the lung and breast [67–70], implying that EGR1 plays a tumor suppressive role in various cancers. This notion is further corroborated by the observation that tumor suppressors such as PTEN and p53 are direct targets of EGR1 [16, 20, 22]. On the other hand, expression of EGR1 is up-regulated in prostate tumors [71–75], implying that the role of EGR1 is tissue-dependent and that its ability to serve both as tumor suppressor and oncoprotein depending on the biological context is an intricate virtue of its functional duality. Importantly, EGR1-responsive genes harbor a rather high degree of DNA sequence variations within their promoters though the physiological and pathological

consequences of such single nucleotide polymorphisms are not fully understood [24–30]. Thus, understanding EGR1-DNA interaction in molecular terms remains an important challenge toward the design of novel therapeutic approaches.

Toward this goal, our work reported here provides the maiden study to quantitatively address the effect of nucleotide substitutions on EGR1-DNA interaction in detailed thermodynamic terms. In particular, we have demonstrated here that the introduction of single nucleotide variations within the ZRE motif can have dramatic effect on EGR1-DNA interaction and that certain positions are much more susceptible to such genetic changes than others. Our findings underscore the fact that sequence variations within the ZRE motif alter protein-DNA interaction primarily via compromising van der Waals and hydrogen bonding contacts. More importantly, nucleotide substitutions seemingly result in differential solvation of protein-DNA complex and, in particular, the water solvent appears to play a key role in modulating protein-DNA thermodynamics.

In short, our study provides a framework for understanding how DNA sequence variations within the target gene promoters may affect the binding of an important transcription factor central to human health and disease.

Acknowledgments

This work was supported by funds from the National Institutes of Health (Grant# R01-GM083897) and the USylvester Braman Family Breast Cancer Institute to AF. CBM is a recipient of a postdoctoral fellowship from the National Institutes of Health (Award# T32-CA119929).

ABBREVIATIONS

CD	Circular dichroism
DB	DNA-binding (domain)
EGR	Early growth response
EGR1	Early growth response (protein) 1
ITC	Isothermal titration calorimetry
LIC	Ligation-independent cloning
MM	Molecular modeling
SEC	Size-exclusion chromatography
TA	Transactivation (domain)
ZF	Zinc finger
ZRE	Zif268 (EGR1) response element

References

1. Marini JC, Levene SD, Crothers DM, Englund PT. Proc Natl Acad Sci U S A. 1982; 79:7664–7668. [PubMed: 16593261]
2. Wu HM, Crothers DM. Nature. 1984; 308:509–13. [PubMed: 6323997]

3. Munteanu MG, Vlahovicek K, Parthasarathy S, Simon I, Pongor S. Trends Biochem Sci. 1998; 23:341–7. [PubMed: 9787640]
4. Pavletich NP, Pabo CO. Science. 1991; 252:809–17. [PubMed: 2028256]
5. Yan SF, Fujita T, Lu J, Okada K, Shan Zou Y, Mackman N, Pinsky DJ, Stern DM. Nat Med. 2000; 6:1355–61. [PubMed: 11100120]
6. Adamson ED, Mercola D. Tumour Biol. 2002; 23:93–102. [PubMed: 12065847]
7. Thiel G, Cibelli G. J Cell Physiol. 2002; 193:287–92. [PubMed: 12384981]
8. Ahmed MM. Curr Cancer Drug Targets. 2004; 4:43–52. [PubMed: 14965266]
9. Bae SK, Bae MH, Ahn MY, Son MJ, Lee YM, Bae MK, Lee OH, Park BC, Kim KW. Cancer Res. 1999; 59:5989–94. [PubMed: 10606246]
10. Liu C, Yao J, de Belle I, Huang RP, Adamson E, Mercola D. J Biol Chem. 1999; 274:4400–11. [PubMed: 9933644]
11. Minc E, de Coppet P, Masson P, Thiery L, Dutertre S, Amor-Gueret M, Jaulin C. J Biol Chem. 1999; 274:503–9. [PubMed: 9867871]
12. Liu C, Yao J, Mercola D, Adamson E. J Biol Chem. 2000; 275:20315–23. [PubMed: 10783396]
13. Svaren J, Ehrig T, Abdulkadir SA, Ehrenguber MU, Watson MA, Milbrandt J. J Biol Chem. 2000; 275:38524–31. [PubMed: 10984481]
14. Bain G, Cravatt CB, Loomans C, Alberola-Ila J, Hedrick SM, Murre C. Nat Immunol. 2001; 2:165–71. [PubMed: 11175815]
15. Guha M, O'Connell MA, Pawlinski R, Hollis A, McGovern P, Yan SF, Stern D, Mackman N. Blood. 2001; 98:1429–39. [PubMed: 11520792]
16. Virolle T, Adamson ED, Baron V, Birle D, Mercola D, Mustelin T, de Belle I. Nat Cell Biol. 2001; 3:1124–8. [PubMed: 11781575]
17. Fu M, Zhang J, Lin Y, Zhu X, Ehrenguber MU, Chen YE. J Biol Chem. 2002; 277:26808–14. [PubMed: 12011097]
18. Fu M, Zhang J, Lin Y, Zhu X, Zhao L, Ahmad M, Ehrenguber MU, Chen YE. Biochem J. 2003; 370:1019–25. [PubMed: 12457461]
19. Renbaum P, Beerli R, Gabai E, Amiel M, Gal M, Ehrenguber MU, Levy-Lahad E. Gene. 2003; 318:113–24. [PubMed: 14585504]
20. Krones-Herzig A, Mittal S, Yule K, Liang H, English C, Urcis R, Soni T, Adamson ED, Mercola D. Cancer Res. 2005; 65:5133–43. [PubMed: 15958557]
21. Kumbink J, Gerlinger M, Johnson JP. J Biol Chem. 2005; 280:42785–93. [PubMed: 16260776]
22. Baron V, Adamson ED, Calogero A, Ragona G, Mercola D. Cancer Gene Ther. 2006; 13:115–24. [PubMed: 16138117]
23. Zagurovskaya M, Shareef MM, Das A, Reeves A, Gupta S, Sudol M, Bedford MT, Prichard J, Mohiuddin M, Ahmed MM. Oncogene. 2009; 28:1121–31. [PubMed: 19137013]
24. Bulyk ML, Huang X, Choo Y, Church GM. Proc Natl Acad Sci U S A. 2001; 98:7158–63. [PubMed: 11404456]
25. Bulyk ML, Johnson PL, Church GM. Nucleic Acids Res. 2002; 30:1255–61. [PubMed: 11861919]
26. Fu M, Zhu X, Zhang J, Liang J, Lin Y, Zhao L, Ehrenguber MU, Chen YE. Gene. 2003; 315:33–41. [PubMed: 14557062]
27. Gonzalez P, Reguero JR, Lozano I, Moris C, Coto E. Int J Immunogenet. 2007; 34:127–30. [PubMed: 17373938]
28. Alvarez V, Gonzalez P, Corao AI, Menendez M, Lahoz CH, Martinez C, Calatayud M, Morales B, Coto E. Alzheimer Dis Assoc Disord. 2008; 22:177–80. [PubMed: 18525291]
29. Geiger EV, Doebering A, Kirchhof A, Lotsch J. Prostaglandins Leukot Essent Fatty Acids. 2009; 80:255–62. [PubMed: 19446447]
30. Kubosaki A, Tomaru Y, Tagami M, Arner E, Miura H, Suzuki T, Suzuki M, Suzuki H, Hayashizaki Y. Genome Biol. 2009; 10:R41. [PubMed: 19374776]
31. Mikles DC, Bhat V, Schuchardt BJ, Deegan BJ, Seldeen KL, McDonald CB, Farooq A. FEBS J. 2013; 280:3669–84. [PubMed: 23718776]

32. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, MR.; Appel, RD.; Bairoch, A. *The Proteomics Protocols Handbook*. Walker, JM., editor. Humana Press; Totowa, New Jersey, USA: 2005. p. 571-607.
33. Cantor CR, Warshaw MM, Shapiro H. *Biopolymers*. 1970; 9:1059–77. [PubMed: 5449435]
34. Wiseman T, Williston S, Brandts JF, Lin LN. *Anal Biochem*. 1989; 179:131–137. [PubMed: 2757186]
35. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. *Annu Rev Biophys Biomol Struct*. 2000; 29:291–325. [PubMed: 10940251]
36. Carson M. *J Appl Crystallogr*. 1991; 24:958–961.
37. Murphy KP, Freire E. *Adv Protein Chem*. 1992; 43:313–61. [PubMed: 1442323]
38. Freire E. *Arch Biochem Biophys*. 1993; 303:181–4. [PubMed: 8512307]
39. Spolar RS, Record JMT. *Science*. 1994; 263:777–784. [PubMed: 8303294]
40. Xie D, Freire E. *Proteins*. 1994; 19:291–301. [PubMed: 7984625]
41. Murphy KP. *Med Res Rev*. 1999; 19:333–9. [PubMed: 10398929]
42. Edgcomb SP, Murphy KP. *Curr Opin Biotechnol*. 2000; 11:62–6. [PubMed: 10679345]
43. Fraczekiewicz R, Braun W. *J Comp Chem*. 1998; 19:319–333.
44. Choo Y, Klug A. *Proc Natl Acad Sci U S A*. 1994; 91:11168–72. [PubMed: 7972028]
45. Desjarlais JR, Berg JM. *Proc Natl Acad Sci U S A*. 1994; 91:11099–103. [PubMed: 7972017]
46. Hamilton TB, Borel F, Romaniuk PJ. *Biochemistry*. 1998; 37:2051–8. [PubMed: 9485332]
47. Wolfe SA, Greisman HA, Ramm EI, Pabo CO. *J Mol Biol*. 1999; 285:1917–34. [PubMed: 9925775]
48. Lumry R, Rajender S. *Biopolymers*. 1970; 9:1125–227. [PubMed: 4918636]
49. Eftink MR, Anusiem AC, Biltonen RL. *Biochemistry*. 1983; 22:3884–96. [PubMed: 6615806]
50. Starikov EB, Norden B. *J Phys Chem B*. 2007; 111:14431–5. [PubMed: 18044870]
51. Olsson TS, Ladbury JE, Pitt WR, Williams MA. *Protein Sci*. 2011; 20:1607–18. [PubMed: 21739503]
52. Chodera JD, Mobley DL. *Annu Rev Biophys*. 2013; 42:121–42. [PubMed: 23654303]
53. Cooper A, Johnson CM, Lakey JH, Nollmann M. *Biophys Chem*. 2001; 93:215–30. [PubMed: 11804727]
54. Foguel D, Silva JL. *Proc Natl Acad Sci U S A*. 1994; 91:8244–7. [PubMed: 8058788]
55. Ladbury JE, Wright JG, Sturtevant JM, Sigler PB. *J Mol Biol*. 1994; 238:669–681. [PubMed: 8182742]
56. Merabet E, Ackers GK. *Biochemistry*. 1995; 34:8554–63. [PubMed: 7612597]
57. Petri V, Hsieh M, Brenowitz M. *Biochemistry*. 1995; 34:9977–84. [PubMed: 7632696]
58. Berger C, Jelesarov I, Bosshard HR. *Biochemistry*. 1996; 35:14984–91. [PubMed: 8942664]
59. Milev S, Bosshard HR, Jelesarov I. *Biochemistry*. 2005; 44:285–93. [PubMed: 15628870]
60. Privalov PL, Dragan AI, Crane-Robinson C. *Trends Biochem Sci*. 2009; 34:464–70. [PubMed: 19726198]
61. Morton CJ, Ladbury JE. *Protein Sci*. 1996; 5:2115–8. [PubMed: 8897612]
62. Bergqvist S, Williams MA, O'Brien R, Ladbury JE. *J Mol Biol*. 2004; 336:829–42. [PubMed: 15095863]
63. Cooper A. *Biophys Chem*. 2005; 115:89–97. [PubMed: 15752588]
64. Chavelas EA, Garcia-Hernandez E. *Biochem J*. 2009; 420:239–47. [PubMed: 19257855]
65. Seldeen KL, McDonald CB, Deegan BJ, Farooq A. *Arch Biochem Biophys*. 2008; 473:48–60. [PubMed: 18316037]
66. Seldeen KL, McDonald CB, Deegan BJ, Farooq A. *Arch Biochem Biophys*. 2008; 480:75–84. [PubMed: 18940179]
67. Levin WJ, Press MF, Gaynor RB, Sukhatme VP, Boone TC, Reissmann PT, Figlin RA, Holmes EC, Souza LM, Slamon DJ. *Oncogene*. 1995; 11:1261–9. [PubMed: 7478546]
68. Huang RP, Fan Y, de Belle I, Niemeyer C, Gottardis MM, Mercola D, Adamson ED. *Int J Cancer*. 1997; 72:102–9. [PubMed: 9212230]

69. Calogero A, Arcella A, De Gregorio G, Porcellini A, Mercola D, Liu C, Lombardi V, Zani M, Giannini G, Gagliardi FM, Caruso R, Gulino A, Frati L, Ragona G. Clin Cancer Res. 2001; 7:2788–96. [PubMed: 11555594]
70. Joslin JM, Fernald AA, Tennant TR, Davis EM, Kogan SC, Anastasi J, Crispino JD, Le Beau MM. Blood. 2007; 110:719–26. [PubMed: 17420284]
71. Thigpen AE, Cala KM, Guileyardo JM, Molberg KH, McConnell JD, Russell DW. J Urol. 1996; 155:975–81. [PubMed: 8583621]
72. Eid MA, Kumar MV, Iczkowski KA, Bostwick DG, Tindall DJ. Cancer Res. 1998; 58:2461–8. [PubMed: 9622090]
73. Baron V, De Gregorio G, Krones-Herzig A, Virolle T, Calogero A, Urcis R, Mercola D. Oncogene. 2003; 22:4194–204. [PubMed: 12833142]
74. Baron V, Duss S, Rhim J, Mercola D. Ann N Y Acad Sci. 2003; 1002:197–216. [PubMed: 14751836]
75. Virolle T, Krones-Herzig A, Baron V, De Gregorio G, Adamson ED, Mercola D. J Biol Chem. 2003; 278:11802–10. [PubMed: 12556466]

Figure 1.

Nucleotide sequence of dsDNA oligos containing the GCGTGGGCG consensus sequence and its single nucleotide variants thereof. (a) In the ZRE motif, the consensus nucleotides are capitalized whilst the flanking nucleotides are shown in small letters, and the three sub-sites that accommodate the zinc fingers of the DB domain of EGR1, namely ZFI (right site), ZFII (middle site) and ZFIII (left site), are marked for clarity. Note that the numbering of various nucleotides with respect to the central nucleotide of the middle site (which is arbitrarily assigned zero) are indicated. (b) G-1 variant motif containing the T→G substitution at the -1 position (colored red and underlined). (c) T+1 variant motif containing the G→T substitution at the +1 position (colored red and underlined).

Figure 2.

Representative ITC isotherms for the binding of DB domain of EGR1 to ZRE consensus motif (a), G-1 variant motif (b), and T +1 variant motif (c). The upper panels show raw ITC data expressed as thermal power with respect to time over the period of titration. In the lower panels, molar heat is expressed as a function of molar ratio of each dsDNA oligo to DB domain. The solid lines in the lower panels represent the fit of data to a one-site model, using the integrated NanoAnalyze software as described previously [34, 31].

Figure 3.

Analysis of the binding of DB domain of EGR1 to variant motifs relative to the ZRE consensus motif in terms of relative binding affinity (K_t), relative enthalpic contribution ($\Delta\Delta H$), relative entropic contribution ($T\Delta\Delta S$) and relative free energy ($\Delta\Delta G$) shown respectively in (a), (b), (c) and (d). K_t is defined as $K_t = K_v/K_c$, where K_v and K_c are respectively the equilibrium dissociation constants of the variant and consensus motifs to the DB domain (Table 1). The thermodynamic terms are defined as $\Delta\Delta H = H_v - H_c$, $T\Delta\Delta S = T S_v - T S_c$ and $\Delta\Delta G = G_v - G_c$, where the subscripts v and c respectively denote the corresponding values observed for the binding of variant and consensus motifs to the DB domain (Table 1). Note that a positive value of $\Delta\Delta H$ is indicative of favorable enthalpic contribution being smaller than that observed for ZRE motif, while a negative value represents a larger favorable enthalpic contribution. Conversely, a positive value of $T\Delta\Delta S$ is indicative of unfavorable entropic

contribution being smaller than that observed for ZRE motif, while a negative value represents a larger unfavorable entropic contribution.

Figure 4.

Structural model of the DB domain of EGR1 in complex with ZRE duplex. Note that the DB domain is comprised of three tandem C2H2-type zinc fingers, designated herein ZFI (olive), ZFII (cyan) and ZFIII (yellow). Each zinc finger is comprised of an α -helix and a double-stranded antiparallel β -sheet that together sandwich a Zn^{2+} divalent ion (omitted for clarity). The DNA backbone is shown in gray and the bases are omitted for clarity, except those located at the -1 and $+1$ positions within the sense strand (colored red and the label underlined) and their counterparts within the antisense strand (colored red). The expanded window shows a close-up view of the sidechain moieties of amino acid residues (colored blue) located within the DB domain that engage in van der Waals contact with DNA bases at the -1 and $+1$ positions (colored red). Note that the dotted lines indicate van der Waals contact between $\text{C}\gamma 2$ -methyl group of T385 and $\text{C}7$ -methyl moiety of T-1 (sense strand), hydrogen bonding between $\text{O}\delta 2$ -hydroxyl oxygen of D381 and $\text{N}4$ -ring nitrogen of C-1 (antisense strand), and hydrogen bonding between $\text{N}\eta 1$ -guanidine nitrogen of R357 and $\text{N}7$ -ring nitrogen of G+1 (sense strand).

Figure 5.

Interdependence of enthalpic (ΔH) and entropic ($T\Delta S$) contributions to the free energy (ΔG) for the binding of DB domain of EGR1 to dsDNA oligos containing the consensus ZRE and variant motifs. (a) ΔH - $T\Delta S$ plot. (b) ΔH - ΔG plot. (c) $T\Delta S$ - ΔG plot. In (a), the red solid line represents linear fit to the data. The coordinates of ZRE motif, the G-1 variant and the T+1 variant are indicated on each plot. Error bars were calculated from at least three independent measurements to one standard deviation.

Figure 6.

Dependence of enthalpic (ΔH) and entropic ($T\Delta S$) contributions to the free energy (ΔG) on the position of nucleotide substitution within the ZRE motif. The changes in various thermodynamic parameters upon the introduction of A (red), T (green), G (blue) and C (purple) at various positions within the consensus motif are color-coded and connected by solid lines for clarity. Error bars were calculated from at least three independent measurements to one standard deviation.

Table 1

Thermodynamic parameters for the binding of DB domain of EGR1 to dsDNA oligos containing the consensus GCGTGGGCG motif (ZRE) and all possible single nucleotide variants thereof

Motif	Sequence	K _d /μM	H/kcal.mol ⁻¹	T S/kcal.mol ⁻¹	G/kcal.mol ⁻¹	Gene Promoter
ZRE	GCGTGGGCG	0.31 ± 0.04	-18.67 ± 0.50	-9.78 ± 0.43	-8.89 ± 0.07	IGF2 growth factor
A-4	ACGTGGGCG	0.47 ± 0.08	-19.67 ± 0.90	-11.02 ± 0.79	-8.65 ± 0.11	BCL2 apoptosis regulator
C-4	CCGTGGGCG	0.97 ± 0.14	-14.85 ± 0.66	-6.64 ± 0.58	-8.21 ± 0.09	CASP8 Caspase
T-4	TCGTGGGCG	1.27 ± 0.28	-16.78 ± 0.28	-8.72 ± 0.42	-8.06 ± 0.13	MECP2 CpG binding protein
A-3	GAGTGGGCG	0.48 ± 0.08	-21.95 ± 0.77	-13.27 ± 0.65	-8.68 ± 0.11	ALOX5 lipoxigenase
G-3	GGGTGGGCG	0.43 ± 0.08	-18.20 ± 0.44	-9.49 ± 0.32	-8.71 ± 0.11	
T-3	GTGTGGGCG	0.65 ± 0.10	-15.97 ± 0.84	-7.52 ± 0.75	-8.45 ± 0.09	NR3C1 nuclear receptor
A-2	GCA ^T TGGGCG	2.39 ± 0.70	-15.30 ± 0.79	-7.61 ± 0.61	-7.69 ± 0.18	TNFα cytokine
C-2	GCC ^T TGGGCG	6.36 ± 0.94	-9.59 ± 0.35	-2.49 ± 0.44	-7.10 ± 0.09	
T-2	GCT ^T TGGGCG	3.96 ± 0.65	-4.60 ± 0.18	+2.78 ± 0.09	-7.38 ± 0.10	TFF1 secretory protein
A-1	GCGA ^T GGGCG	1.59 ± 0.22	-20.04 ± 0.44	-12.12 ± 0.36	-7.92 ± 0.08	PRCC mitotic protein
C-1	GCGC ^T GGGCG	1.41 ± 0.19	-17.87 ± 0.53	-9.88 ± 0.45	-7.99 ± 0.08	EGFR receptor tyrosine kinase
G-1	GCGG ^T GGGCG	1.10 ± 0.21	-28.07 ± 0.49	-19.93 ± 0.38	-8.14 ± 0.11	MMP14 matrix metalloproteinase
A0	GCGT ^A GGCG	0.63 ± 0.23	-15.42 ± 0.30	-6.93 ± 0.08	-8.49 ± 0.22	p53 tumor suppressor
C0	GCGT ^C GGCG	1.91 ± 0.55	-12.07 ± 0.40	-4.25 ± 0.23	-7.82 ± 0.17	

Motif	Sequence	K _d /μM	H/kcal.mol ⁻¹	T S/kcal.mol ⁻¹	G/kcal.mol ⁻¹	Gene Promoter
T0	GCGT T GGCG	0.94 ± 0.14	-12.58 ± 0.23	-4.35 ± 0.15	-8.23 ± 0.09	
A+1	GCGT G AGCG	3.38 ± 0.41	-18.35 ± 0.83	-10.87 ± 0.90	-7.48 ± 0.07	LIF cytokine
C+1	GCGT G CGCG	7.26 ± 1.42	-10.88 ± 0.41	-3.86 ± 0.29	-7.02 ± 0.12	ID3 DNA-binding inhibitor
T+1	GCGT T GCG	7.46 ± 1.41	-22.34 ± 1.10	-15.34 ± 0.99	-7.01 ± 0.11	
A+2	GCGT G ACG	5.82 ± 1.27	-18.07 ± 0.56	-10.91 ± 0.43	-7.16 ± 0.13	CAPS calcium-binding protein
C+2	GCGT G CGG	5.01 ± 0.78	-16.67 ± 0.42	-9.42 ± 0.33	-7.24 ± 0.09	TJP1 tight junction protein
T+2	GCGT G TG	1.66 ± 0.44	-14.65 ± 0.84	-6.75 ± 0.68	-7.90 ± 0.16	PTEN protein phosphatase
A+3	GCGT G GGAG	0.85 ± 0.09	-16.40 ± 0.69	-8.11 ± 0.63	-8.29 ± 0.06	BAD apoptotic regulator
G+3	GCGT G GGG	1.02 ± 0.19	-17.27 ± 0.85	-9.08 ± 0.96	-8.19 ± 0.11	APEX1 endonuclease
T+3	GCGT G GGT G	1.54 ± 0.36	-14.90 ± 0.69	-6.95 ± 0.55	-7.95 ± 0.14	VEGF growth factor
A+4	GCGT G GG C A	1.07 ± 0.32	-10.69 ± 0.63	-2.52 ± 0.45	-8.17 ± 0.18	RBL1 retinoblastoma protein
C+4	GCGT G GG C C	1.27 ± 0.32	-11.95 ± 0.55	-3.89 ± 0.40	-8.06 ± 0.15	ID11 cytokine
T+4	GCGT G GG C T	1.38 ± 0.35	-6.51 ± 0.17	+1.50 ± 0.01	-8.01 ± 0.15	BAX apoptotic regulator

Note that the DNA sequence for the GCGTGGGCG motif and its single nucleotide variants corresponds to the sense strand only and nucleotides flanking these motifs have been omitted (see Figure 1 for details). The substituted nucleotide relative to the consensus GCGTGGGCG motif is colored red and underlined. One example (where known) of an EGRI-responsive gene promoter that contains at least one of the substitutions within the corresponding variant of ZRE motif, identified using GeneCards online server located @ <http://www.genecards.org>, is provided for physiological relevance. Binding stoichiometries of DB domain to dsDNA oligos were generally within ±10% of unity. Errors were calculated from at least three independent measurements and are given to one standard deviation.

Effect of solvation on the binding of DB domain of EGR1 to dsDNA oligos containing the consensus GCGTGGGCG motif (ZRE) and all possible single nucleotide variants thereof

Table 2

Motif	Sequence	– $H_{60}/\text{kcal.mol}^{-1}$	– $C_p/\text{cal.mol}^{-1}.\text{K}^{-1}$	– $\text{SASA}_{\text{polar}}/\text{\AA}^2$	– $\text{SASA}_{\text{apolar}}/\text{\AA}^2$	– $\text{SASA}_{\text{total}}/\text{\AA}^2$
ZRE	GCGTGGGCG	30.71 ± 0.92	341 ± 24	1403 ± 52	1570 ± 81	2972 ± 132
A-4	ACGTGGGCG	35.29 ± 0.95	442 ± 23	1647 ± 51	1934 ± 80	3581 ± 131
C-4	CCGTGGGCG	29.03 ± 2.21	399 ± 57	1380 ± 123	1684 ± 197	3064 ± 321
T-4	TCGTGGGCG	29.15 ± 2.22	350 ± 65	1350 ± 130	1558 ± 220	2907 ± 349
A-3	GAGTGGGCG	36.29 ± 0.96	388 ± 14	1646 ± 47	1814 ± 59	3460 ± 105
G-3	GGGTGGGCG	33.18 ± 0.44	424 ± 12	1555 ± 26	1842 ± 41	3397 ± 66
T-3	GTTGGGCG	33.32 ± 0.58	504 ± 5	1616 ± 25	2053 ± 26	3668 ± 51
A-2	GCA TGGGCG	26.66 ± 0.33	332 ± 9	1243 ± 19	1455 ± 32	2697 ± 50
C-2	GCCTGGGCG	21.31 ± 1.01	321 ± 25	1034 ± 56	1312 ± 87	2345 ± 143
T-2	GCTTGGGCG	15.80 ± 0.82	308 ± 17	816 ± 43	1156 ± 63	1971 ± 105
A-1	GCGAGGGCG	32.49 ± 0.27	355 ± 7	1487 ± 15	1648 ± 25	3135 ± 40
C-1	GCGCGGGCG	35.32 ± 0.90	499 ± 17	1689 ± 46	2085 ± 64	3773 ± 110
G-1	GCGGGGGCG	41.69 ± 1.15	400 ± 29	1860 ± 64	1964 ± 102	3824 ± 166
A0	GCGTAGGCG	33.22 ± 0.41	510 ± 10	1617 ± 23	2068 ± 35	3685 ± 57
C0	GCGTCGGCG	24.33 ± 1.29	357 ± 33	1173 ± 71	1471 ± 114	2643 ± 185

Motif	Sequence	$-H_{60}/\text{kcal.mol}^{-1}$	$-C_p/\text{cal.mol}^{-1}\text{K}^{-1}$	$-SASA_{\text{polar}}/\text{\AA}^2$	$-SASA_{\text{apolar}}/\text{\AA}^2$	$-SASA_{\text{total}}/\text{\AA}^2$
T0	GCGT T GGCG	30.55 ± 1.05	499 ± 28	1509 ± 59	1982 ± 95	3491 ± 155
A+1	GCGT G AGCG	31.03 ± 1.81	350 ± 44	1421 ± 99	1600 ± 153	3021 ± 252
C+1	GCGT G CGCG	26.17 ± 1.78	420 ± 52	1287 ± 104	1677 ± 177	2964 ± 281
T+1	GCGT G TGCG	34.45 ± 1.49	351 ± 43	1551 ± 86	1676 ± 144	3227 ± 231
A+2	GCGT G ACG	29.59 ± 0.47	315 ± 12	1342 ± 26	1477 ± 42	2818 ± 68
C+2	GCGT G CCG	29.76 ± 0.67	366 ± 24	1384 ± 42	1613 ± 78	2997 ± 120
T+2	GCGT G TCG	26.67 ± 1.39	340 ± 33	1249 ± 75	1479 ± 114	2728 ± 189
A+3	GCGT G GGAG	33.63 ± 1.31	480 ± 23	1611 ± 65	1998 ± 87	3609 ± 153
G+3	GCGT G GGGG	26.70 ± 0.99	269 ± 37	1200 ± 63	1291 ± 118	2491 ± 182
T+3	GCGT G GGTG	29.62 ± 0.46	414 ± 15	1413 ± 29	1735 ± 50	3148 ± 78
A+4	GCGT G GGCA	24.94 ± 0.99	402 ± 34	1228 ± 62	1603 ± 111	2830 ± 173
C+4	GCGT G GGCC	22.87 ± 0.66	332 ± 9	1100 ± 30	1373 ± 37	2473 ± 67
T+4	GCGT G GGCT	17.53 ± 1.30	312 ± 35	870 ± 94	1197 ± 132	2067 ± 226

Note that the DNA sequence for the GCGTGGCG motif and its single nucleotide variants corresponds to the sense strand only and nucleotides flanking these motifs have been omitted (see Figure 1 for details). The substituted nucleotide relative to the consensus GCGTGGCG motif is colored red and underlined. Errors were calculated from at least three independent measurements and are given to one standard deviation.