Local Conformational Changes in the DNA Interfaces of Proteins

Tomoko Sunami, Hidetoshi Kono *

Molecular Modeling and Simulation Group, Quantum Beam Science Directorate, Japan Atomic Energy Agency, Kizugawa, Kyoto, Japan

Abstract

When a protein binds to DNA, a conformational change is often induced so that the protein will fit into the DNA structure. Therefore, quantitative analyses were conducted to understand the conformational changes in proteins. The results showed that conformational changes in DNA interfaces are more frequent than in non-interfaces, and DNA interfaces have more conformational variations in the DNA-free form. As expected, the former indicates that interaction with DNA has some influence on protein structure. The latter suggests that the intrinsic conformational flexibility of DNA interfaces is important for adjusting their conformation for DNA. The amino acid propensities of the conformationally changed regions in DNA interfaces indicate that hydrophilic residues are preferred over the amino acids that appear in the conformationally unchanged regions. This trend is true for disordered regions, suggesting again that intrinsic flexibility is of importance not only for DNA binding but also for interactions with other molecules. These results demonstrate that fragments destined to be DNA interfaces have an intrinsic flexibility and are composed of amino acids with the capability of binding to DNA. This information suggests that the prediction of DNA binding sites may be improved by the integration of amino acid preference for DNA and one for disordered regions.

Introduction

Protein–DNA interaction plays an essential role in many cellular functions such as transcription, replication, recombination, and DNA packaging. To understand the recognition mechanisms of individual DNA binding proteins, the protein structures of DNA-bound as well as DNA-free forms have been analyzed [1,2,3,4,5]. It has been reported that flexible regions undergo conformational changes in order to recognize specific DNA targets [4,5,6,7,8,9,10]. For example, the β2/β3 connecting loop of the papillomavirus E2 protein, which is unstructured in the free form, adopts a β-hairpin conformation in order to form electrostatic contacts with DNA backbone phosphates in the complex form [4,6]. The conformational change in the loop has also been observed in molecular dynamics simulations [7,8]. Another example of conformational change was observed in the linker region of MATα2 [5]. In this case, two independent copies of the complex were found in the asymmetric unit. The flexible linker [9] in one copy of MATα2 adopted an α-helix structure and the other adopted a β-strand structure. The sequence of the region is coined as a chameleon sequence. This conformational transition at the sequence is thought to be important for DNA recognition [10].

The sequence characteristics of such proteins have also been examined. Dunker and other groups developed methods to predict the intrinsically disordered region of proteins on the basis of sequence analysis, nuclear magnetic resonance, and circular dichroism spectroscopic data [11,12]. Such regions are thought to undergo a disordered–ordered transition of the conformations when they interact with a binding partner [11]. The genome-wide application of these methods indicated that transcription factors, especially those in eukaryotes, have a higher amount of intrinsically disordered regions [13,14]. The proposed role of the regions is to facilitate DNA searching and modulate the specificity and affinity to DNA [15]. Although the conformational change in the flexible region in DNA binding proteins has been well recognized, comprehensive analyses of the local structural rearrangements of proteins upon DNA binding have not been conducted yet.

To assess the structural rearrangement, the classification of the 3D geometries of local protein structures is necessary. Historically, Pauling first proposed the idea that protein structures could be represented as strings of secondary structures [16], and since then, secondary structures have often been used to compare protein structures [17,18]. However, secondary structures are too coarse to detect subtle local conformations because they only focus on the arrangement of the hydrogen bonding partners of the backbone atoms. The use of structural alphabets was then proposed to more precisely describe local structures, where the alphabets are assigned to certain local conformations [19]. Structural alphabets have been reported to classify protein structures more precisely than secondary structures [20,21] and have been applied to structure prediction [22,23], 3D structure comparison [24,25], motif searches [26], protein–protein interaction analysis [27,28], and de novo protein design [29].

In this study, quantitative analysis of conformational changes in DNA binding proteins using structural alphabets [30] was performed. Using sets of proteins whose structures were solved for both the DNA-free and DNA-bound forms, it was found that
DNA interfaces have higher conformational flexibility than non-interfaces. It was also found that conformationally changed regions in DNA interfaces have a high amount of glycine, proline, and the hydrophilic residues that have previously been found in intrinsically disordered regions [11,12,31]. This result indicates that fragments of DNA interfaces are composed of amino acids that have high flexibility and DNA binding capability.

Materials and Methods

Data Preparation for DNA-bound and DNA-free Forms

Non-redundant pairs of crystal structures of DNA-free and DNA-bound forms were prepared as follows. A flow chart of the data preparation and a schematic diagram of the reduction of the dataset redundancy are given in Figure 1.

To prepare the dataset for the DNA-bound forms, the Protein Data Bank (PDB) (December 2010 version [32]) was searched for all DNA complexes with better than 3 Å resolution. The proteins that were co-crystallized with ss-DNA, Z-DNA, and RNA were then discarded. Antibodies, artificial DNA binding proteins, and a structure of a trp repressor that was crystallized with a high concentration of isopropanol (pdbID: 1mi7) were further excluded. The PDB was then searched for the DNA-free forms of the DNA-bound proteins (>90% sequence identity) with a resolution better than 3 Å. To reduce the dataset redundancy, the selected proteins with a sequence identity of 30% were clustered using Blastclust [33]. Subclusters with a sequence identity of 90% were then made within each of the clusters. In each cluster, the representative subcluster was determined to be the one that contained the largest number of protein chains among all subclusters. Finally, 126 representative cluster pairs of DNA-bound forms and DNA-free forms were obtained. Hereafter, these representatives are referred to as DBbound and DBfree, respectively.

To evaluate the conformational variation in the DNA free forms of the proteins, another dataset was prepared. Clusters that had more than two protein chains were extracted from DBfree. This dataset is referred to as DBfree(>2), and contained 86 clusters. The members of DBbound, DBfree, and DBfree(>2) are listed in Table S1.

Assignment of 11 Structural Alphabets

A library composed of 10 4-residue-long fragments that were developed by Kolodony et al. [30] and one fragment we introduced in this study was used to describe the protein structures. For describing the fragments, alphabets A to J were assigned (Fig. 2(a)) and their conformations are shown in Table S2. Hereafter, we call them structural alphabets. In addition to the original 10 fragments, one “Y” code was introduced to describe a fragment for which any of the Cα atoms were not determined in the crystal structure. Such fragments are thought to acquire multi-conformations in a crystal. To each structural fragment, the best-matched alphabet in terms of the root mean square deviation of the Cα atoms (cRMS) was assigned. The fragments corresponding to 5% outliers in the cRMS distributions were discarded from the analysis. In addition, fragments were excluded from the analysis if the sequences of the corresponding fragments were not identical.

![Figure 1. Data set preparation.](https://example.com/figure1)

(a) Workflow for obtaining DBbound, DBfree, and DBfree(>2). (b) Schematic diagram for the preparation of a set of non-redundant clusters.

doi:10.1371/journal.pone.0056080.g001
among the proteins of the same cluster. To intuitively catch conformational feature of fragments, the ten alphabet structures were further classified into three conformations, extended, loop-like and helix-like conformations based on the cRMS (Fig. 2(b)).

**DNA Interface/non-interface Assignment**

A DNA interface residue was one that was exposed to both the solvent as well as the DNA. To determine such residues, the solvent accessible surface area (ASA) of the proteins in the DNA-bound form was first calculated after removing the bound DNA using the ASC program [34]. The relative ASA for each residue was then calculated as the ratio of the surface area of a residue in the protein structure to that of a residue in a Gly-X-Gly tri-peptide having the trans form. The surface residues were defined as those with a relative ASA of more than 20%. The ASA on the protein structure in the presence of the bound DNA was also calculated. If the ASA of the residue exposed to the solvent was different between the DNA-bound and unbound structures, the residue was considered to be within the DNA interface. In addition, if at least one residue of a fragment was judged to be within the DNA interface, the fragment was considered a DNA interface and the remaining fragments were regarded as DNA non-interfaces.

**Conformational Changes Upon DNA Binding**

Conformational changes were considered to be the differences in the structural alphabets of the fragments in the DNA-free and the corresponding DNA-bound forms. The probability of the conformational change of fragment \( I \) from alphabet \( i \) in DNA-free forms \( \left( w_{chg,I}^{j} \right) \) was calculated as follows:

\[
w_{chg,I}^{j} = \frac{n_{i}^{j} p_{l}^{j}}{\sum_{j=1}^{n_{i}^{j}} p_{l}^{j}}
\]

where \( n_{i}^{j} \) is the number of alphabet \( i \) assigned to fragment \( I \) in DNA-free forms. The variable \( p_{l}^{j} \) is the probability of fragment \( I \) having alphabet \( i \) in the DNA-bound form and is given by

\[
p_{l}^{j} = \frac{N_{i}^{j}}{\sum_{j=1}^{N_{i}^{j}} p_{l}^{j}}
\]

where \( N_{i}^{j} \) is the number of alphabet \( i \) assigned to fragment \( I \) in the DNA-bound forms.

The frequencies of the conformational change in the DNA interfaces \( f_{chg, interface}^{j} \) and non-interfaces \( f_{chg, non-interface}^{j} \) of alphabet \( i \) were calculated as follows:

\[
f_{chg, region}^{j} = \frac{\sum_{region} w_{chg,i}^{j}}{\sum_{region} (n_{i}^{j}/\sum_{j=1}^{n_{i}^{j}} n_{i}^{j})}
\]

where “region” denotes DNA interface or DNA non-interface hereafter.

The alphabet propensity, or the ratio of the frequencies of the conformational changes in the DNA interfaces and non-interfaces \( R_{chg,interface/non-interface}^{j} \) is given by

\[
R_{chg,interface/non-interface}^{j} = \frac{f_{chg,interface}^{j}}{f_{chg,non-interface}^{j}}
\]

The frequencies of the conformational changes upon DNA binding for any alphabet \( f_{chg, interface} \) for interfaces and \( f_{chg, non-interface} \) for non-interfaces.

---

Figure 2. Structure of the eleven fragments. (a) Ten fragments (assigned to A to J) of the 4-residue-long structural alphabet library and a fragment with missing coordinates (assigned to Y). (b) Similarity of the ten alphabet structures calculated using cRMS as the distance with the Fitch in Phylip package. Fragment structures were drawn with Pymol [42]. doi:10.1371/journal.pone.0056080.g002
for non-interfaces) were defined as

\[ f_{\text{region}}^{\text{var}} = \frac{\sum_{\text{lregion}} \sum_{i=1}^{11} w_{\text{var},i}^{l}}{L_{\text{region}}} \]

where \( L_{\text{region}} \) is the number of fragments in region.

**Conformational Variations in the DNA-free Forms**

To determine the conformational variations in the DNA-free forms, the intrinsic conformational variation, which can be observed as alphabet variations within the same fragments obtained from different crystal structures, was considered. The expected alphabet variation of fragment \( l \) (\( w_{\text{var},i}^{l} \)) was calculated using the following equation:

\[ w_{\text{var},i}^{l} = \frac{n_{i}^{l}}{\sum_{j=1}^{n_{f}} n_{j}^{l}} \]

where \( q_{i}^{l} \) is defined as the probability that two members randomly selected from a set of fragment \( l \) will both have alphabet \( i \) and is given as

\[ q_{i}^{l} = \frac{n_{i}^{l} - 1}{\sum_{j=1}^{n_{f}} n_{j}^{l}} = 1 - \frac{1}{q_{i}^{l}}. \]

For the disordered conformation “X,” \( q_{i}^{X} \) was defined as 1.

The frequencies of the conformational variation in DNA interfaces (\( f_{\text{interface}}^{\text{var}} \)) and non-interfaces (\( f_{\text{non-interface}}^{\text{var}} \)) of alphabet \( i \) were calculated as:

\[ f_{\text{region}}^{\text{var},i} = \frac{\sum_{\text{lregion}} w_{\text{var},i}^{l}}{\sum_{\text{lregion}} (n_{i}/\sum_{j=1}^{n_{f}} n_{j}^{l})}. \]

The alphabet propensity of the conformational variation in the DNA interfaces to that in non-interfaces (\( R_{\text{var},i}^{\text{interface/non-interface}} \)) was defined as

\[ R_{\text{var},i}^{\text{interface/non-interface}} = \frac{f_{\text{interface}}^{\text{var},i}}{f_{\text{non-interface}}^{\text{var},i}}. \]

The frequency of the conformational variation for any alphabet (\( f_{\text{interface}}^{\text{var}} \) for interfaces and \( f_{\text{non-interface}}^{\text{var}} \) for non-interfaces) was defined as

\[ f_{\text{region}}^{\text{var}} = \frac{\sum_{\text{lregion}} \sum_{i=1}^{11} w_{\text{var},i}^{l}}{L_{\text{region}}} \]

where \( L_{\text{region}} \) is the number of fragments in region.

**Propensity Calculations**

To measure the relative differences in a pair of frequencies, propensities for various pairs of frequencies were calculated. Here, we describe, for example, how residue propensities are calculated. The frequencies of each amino acid \( r \) in the conformationally changed fragments were calculated for DNA interfaces (\( f_{\text{res=r}}^{\text{interface}} \)) and non-interfaces (\( f_{\text{res=r}}^{\text{non-interface}} \)) with the following equation:

\[ f_{\text{region}}^{\text{res=r}} = \frac{\sum_{\text{lregion}} \sum_{i=1}^{n_{f}} (1 - \sum_{j=1}^{n_{f}} w_{\text{chg},i}^{l}) \delta(r^{l} - r)}{\sum_{\text{lregion}} \sum_{i=1}^{11} \sum_{j=1}^{n_{f}} w_{\text{chg},i}^{l} \delta(r^{l} - r)} \]

where \( r^{l} \) is the \( i \)th amino acid residue of fragment \( l \). The function \( \delta(x) \) is 1 if \( x = 0 \) and 0 otherwise.

The frequencies of each amino acid \( r \) in the conformationally unchanged fragments were calculated for the DNA interfaces (\( f_{\text{res=r}}^{\text{interface}} \)) and the surface (\( f_{\text{res=r}}^{\text{non-interface}} \)) with the following equation:

\[ f_{\text{region}}^{\text{res=r}} = \frac{\sum_{\text{lregion}} \sum_{i=1}^{n_{f}} (1 - \sum_{j=1}^{n_{f}} w_{\text{chg},i}^{l}) \delta(r^{l} - r)}{\sum_{\text{lregion}} \sum_{i=1}^{11} \sum_{j=1}^{n_{f}} w_{\text{chg},i}^{l} \delta(r^{l} - r)} \]

The residue propensities for the DNA interfaces (\( R_{\text{res=r}}^{\text{chg/unchg/interface}} \)) and non-interfaces (\( R_{\text{res=r}}^{\text{chg/unchg/non-interface}} \)) were defined as

\[ R_{\text{res=r}}^{\text{chg/unchg/interface}} = \frac{f_{\text{interface}}^{\text{res=r}}}{f_{\text{res=r}}} \]

and

\[ R_{\text{res=r}}^{\text{chg/unchg/non-interface}} = \frac{f_{\text{non-interface}}^{\text{res=r}}}{f_{\text{res=r}}} \]

In addition, to measure the importance of each DNA interface residue in the conformationally changed and unchanged fragments, the interface-residue propensities for the conformationally changed fragments (\( R_{\text{res=r}}^{\text{interface/non-interface/unchg}} \)) and conformationally unchanged fragments (\( R_{\text{res=r}}^{\text{interface/non-interface/unchg}} \)) were defined as

\[ R_{\text{res=r}}^{\text{interface/non-interface/unchg}} = \frac{f_{\text{interface}}^{\text{res=r}}}{f_{\text{res=r}}} \]

and

\[ R_{\text{res=r}}^{\text{interface/non-interface/unchg}} = \frac{f_{\text{interface}}^{\text{res=r}}}{f_{\text{res=r}}} \]
## Table 1. Database composition.

| Cluster ID | Representative PDB ID(chain)* in DBfree | Representative PDB ID(chain) * in DBbound | Molecular name** | Used for DBfree(2) *** |
|------------|----------------------------------------|------------------------------------------|-----------------|------------------------|
| 1          | 1aqi(A) 2ih2(A)                         | Modification methylase TaqI              |                 | ✓                      |
| 2          | 1aro(P) 1cez(A)                         | Bacteriophage T7 RNA polymerase          |                 | ✓                      |
| 3          | 1ax3(A) 1sv5(A)                         | Type II restriction enzyme EcoRV          |                 | ✓                      |
| 4          | 1b24(A) 2v57(A)                         | Homing endonuclease I-Dmol               |                 | ✓                      |
| 5          | 1bam(A) 3bam(A)                         | Restriction endonuclease BamHI           |                 | ✓                      |
| 6          | 1baz(A) 1bdt(A)                         | Gene-regulating protein Arc              |                 | ✓                      |
| 7          | 1bjt(A) 3hj4(A)                         | DNA topoisomerase 2                      |                 | ✓                      |
| 8          | 1bm9(A) 2f4k(A)                         | Replication termination protein          |                 | ✓                      |
| 9          | 1bpe(A) 2fmp(A)                         | DNA polymerase β                        |                 | ✓                      |
| 10         | 1ci4(A) 2bx7(A)                         | Barrier-to-autointegration factor        |                 | ✓                      |
| 11         | 1c6(B) 1h99(A)                          | CAAT/enhancer binding protein β          |                 | ✓                      |
| 12         | 1cmb(A) 1mjo(A)                         | Methionine repressor                    |                 | ✓                      |
| 13         | 1ceq(A) 1h9d(A)                         | Core-binding factor α subunit1           |                 | ✓                      |
| 14         | 1es8(A) 1dfm(A)                         | Endonuclease BgIII                      |                 | ✓                      |
| 15         | 1eto(A) 3jih(A)                         | DNA-binding protein fis                  |                 | ✓                      |
| 16         | 1ev7(A) 1iw(A)                          | Type II Restriction enzyme NAEI          |                 | ✓                      |
| 17         | 1evx(A) 1a73(A)                         | INTRON 3 (I-Ppo) ENCODED ENDONUCLEASE    |                 | ✓                      |
| 18         | 1f9f(A) 1j4(A)                          | Regulatory protein E2                    |                 | ✓                      |
| 19         | 1fc3(A) 1lq1(A)                         | Stage 0 sporulation protein A            |                 | ✓                      |
| 20         | 1fr2(B) 1v15(A)                         | Colicin E9                               |                 | ✓                      |
| 21         | 1fvi(A) 2q2t(A)                         | Chlorella virus DNA ligase               |                 | ✓                      |
| 22         | 1fx7(A) 1uB8(A)                         | Iron-dependent repressor IdeR            |                 | ✓                      |
| 23         | 1gxo(A) 1gxp(A)                         | Phosphate regulation transcriptional regulatory protein |                 | ✓                      |
| 24         | 1hm(A) 2c7p(A)                          | Modification methylase HhaI              |                 | ✓                      |
| 25         | 1hw5(A) 1zrf(A)                         | Catabolite gene activator               |                 | ✓                      |
| 26         | 1i7(A) 3nae(A)                          | DNA polymerase                          |                 | ✓                      |
| 27         | 1i77(A) 3dsd(B)                         | DNA double-strand break repair protein Mre11 |                 | ✓                      |
| 28         | 1ikn(A) 2saml(A)                        | Transcription factor NF-xB p65           |                 | ✓                      |
| 29         | 1jbg(A) 1r8d(A)                         | Transcription activator MtaN             |                 | ✓                      |
| 30         | 1jeq(A) 1jey(A)                         | Ku70                                     |                 | ✓                      |
| 31         | 1jey1(A) 1jey(B)                        | Ku80                                     |                 | ✓                      |
| 32         | 1jg7(A) 1m5r(A)                         | DNA β-glucosyltransferase               |                 | ✓                      |
| 33         | 1jh8(A) 1tro(A)                         | Trp repressor                           |                 | ✓                      |
| 34         | 1jih(A) 3mfl(A)                         | DNA polymerase η                        |                 | ✓                      |
| 35         | 1jih(A) 2bopi(A)                        | E2                                      |                 | ✓                      |
| 36         | 1jye(A) 1efa(A)                         | Lac repressor                           |                 | ✓                      |
| 37         | 1k0z(A) 3pvi(A)                         | PvuII endonuclease                      |                 | ✓                      |
| 38         | 1ko9(A) 1m3q(A)                         | 8-Oxoguanine DNA glycosylase            |                 | ✓                      |
| 39         | 1ku3(A) 1rio(H)                         | Sigma factor SigA                       |                 | ✓                      |
| 40         | 1lij(A) 1xpx(A)                         | Protein prospero                        |                 | ✓                      |
| 41         | 1mmi(A) 3fsi(A)                         | MMLV Reverse transcriptase domain        |                 | ✓                      |
| 42         | 1mp3(A) 3cw7(A)                         | DNA-3-methyladenine glycosylase 2        |                 | ✓                      |
| 43         | 1mug(A) 1mwi(A)                         | G/U mismatch-specific DNA glycosylase    |                 | ✓                      |
| 44         | 1okr(A) 1sax(A)                         | Methicillin resistance regulatory protein mecl |                 | ✓                      |
| 45         | 1ou0(A) 1oupl(A)                        | Vibrio vulnificus nuclease              |                 | ✓                      |
| 46         | 1owl(A) 1zte(A)                         | Deoxyribodipyrimidinyl photo lyase       |                 | ✓                      |
| 47         | 1p77(A) 2hdh(A)                         | Engreiled homeodomain                   |                 | ✓                      |
| 48         | 1q5s(A) 1yf3(A)                         | DNA adenine methylase                   |                 | ✓                      |
| 49         | 1q3b(A) 1k3x(A)                         | Endonuclease VIII                       |                 | ✓                      |
| Cluster ID | Representative PDB ID(chain)^* in DB_free | Representative PDB ID(chain)^* in DB_bound | Molecular name** | Used for DB_free(2) *** |
|------------|------------------------------------------|---------------------------------------------|------------------|-------------------------|
| 50         | 1q8i(A)                                  | 3ks9(A)                                     | DNA polymerase II | ✓                       |
| 51         | 1qht(A)                                  | 2vwj(A)                                     | Thermococcus gorgonarius DNA polymerase | ✓                       |
| 52         | 1tqw(A)                                  | 2no9(A)                                     | Endonuclease 4    | ✓                       |
| 53         | 1r69(A)                                  | 1per(L)                                     | 434 repressor     | ✓                       |
| 54         | 1sdo(A)                                  | 2p0j(A)                                     | BstYI            | ✓                       |
| 55         | 1tzy(A)                                  | 1ks5(C)                                     | Histone H2A.1    | ✓                       |
| 56         | 1tzy(B)                                  | 1ks5(D)                                     | Histone H2B.2    | ✓                       |
| 57         | 1tzy(C)                                  | 1ks5(A)                                     | Histone H3       | ✓                       |
| 58         | 1tzy(D)                                  | 1ks5(B)                                     | Histone H4       | ✓                       |
| 59         | 1vhi(A)                                  | 1b3t(A)                                     | Nuclear protein EBNA1 | ✓                       |
| 60         | 1vok(A)                                  | 1qna(A)                                     | Transcription initiation factor TFIID-1 | ✓                       |
| 61         | 1vs(A)                                  | 1odgj(A)                                    | DNA mismatch endonuclease | ✓                       |
| 62         | 1w9h(A)                                  | 2w42(A)                                     | Archaeal Piwi protein | ✓                       |
| 63         | 1wtld(A)                                 | 1wte(A)                                     | EcoO109IR        | ✓                       |
| 64         | 1xhw(A)                                  | 2pyj(A)                                     | phi29 DNA polymerase | ✓                       |
| 65         | 1xsl(A)                                  | 1ylz(A)                                     | DNA α-glucosyltransferase | ✓                       |
| 66         | 1xwl(A)                                  | 2hvh(A)                                     | DNA Polymerase I  | ✓                       |
| 67         | 1ytm(A)                                  | 2fkc(A)                                     | R.HinP11 restriction endonuclease | ✓                       |
| 68         | 1z91(A)                                  | 1z9c(A)                                     | Organic hydroperoxide resistance transcriptional regulator | ✓                       |
| 69         | 1zbf(A)                                  | 3ey1(A)                                     | Ribonuclease H    | ✓                       |
| 70         | 2a4o(B)                                  | 2dnj(A)                                     | Deoxyribonuclease I | ✓                       |
| 71         | 2a6m(A)                                  | 2vih(A)                                     | Transposase ORFA  | ✓                       |
| 72         | 2aud(A)                                  | 2gig(A)                                     | Type II restriction enzyme HincII | ✓                       |
| 73         | 2bnk(A)                                  | 2c5r(A)                                     | Early protein p16.7 | ✓                       |
| 74         | 2cxx(A)                                  | 2qtbj(A)                                    | Telomere binding protein TBP1 | ✓                       |
| 75         | 2cpg(A)                                  | 1b01(A)                                     | Transcriptional repressor CopG | ✓                       |
| 76         | 2d3y(A)                                  | 2dpe(A)                                     | Uracil-DNA glycosylase | ✓                       |
| 77         | 2dt5(A)                                  | 3ikt(A)                                     | Redox-sensing transcriptional repressor rex | ✓                       |
| 78         | 2end(A)                                  | 2fccc(A)                                    | Endonuclease V    | ✓                       |
| 79         | 2f4q(A)                                  | 3m4a(A)                                     | Deinococcus radiodurans Type IB DNA topoisomerases | ✓                       |
| 80         | 2flp(A)                                  | 2flo(A)                                     | Late genes activator | ✓                       |
| 81         | 2fok(A)                                  | 1fok(A)                                     | FokI restriction endonuclease | ✓                       |
| 82         | 2fhu(A)                                  | 1fzp(B)                                     | Staphylococcal accessory regulator A | ✓                       |
| 83         | 2fu(A)                                   | 2tit(A)                                     | Large T antigen   | ✓                       |
| 84         | 2gpe(A)                                  | 2bhf(A)                                     | Bifunctional protein putA | ✓                       |
| 85         | 2gpx(A)                                  | 3gfl(A)                                     | ST1710           | ✓                       |
| 86         | 2hts(A)                                  | 3hts(B)                                     | Kluyveromyces lactis heat shock transcription factor | ✓                       |
| 87         | 2iru(A)                                  | 2r9l(A)                                     | Polymerase Domain from Mycobacterium tuberculosis Ligase D | ✓                       |
| 88         | 2nov(A)                                  | 3k9f(A)                                     | DNA topoisomerase 4 subunit A | ✓                       |
| 89         | 2oa9(A)                                  | 2oa9(A)                                     | RMval            | ✓                       |
| 90         | 2o9d(A)                                  | 2o9d(B)                                     | R.BcnI           | ✓                       |
| 91         | 2o9k(A)                                  | 2o9k(A)                                     | 3-Methyladenine DNA glycosylase I, constitutive | ✓                       |
| 92         | 2ore(D)                                  | 2g1p(A)                                     | DNA adenine methylase | ✓                       |
| 93         | 2p5k(A)                                  | 2p5l(C)                                     | Arginine repressor | ✓                       |
| 94         | 2pof(A)                                  | 3c2p(A)                                     | Virion RNA polymerase | ✓                       |
| 95         | 2qfl(A)                                  | 2qfl(A)                                     | DNA repair protein RAD4 | ✓                       |
| 96         | 2rfl(A)                                  | 1jj4(A)                                     | DNA polymerase IV (family Y) | ✓                       |
Table 1. Cont.

| Cluster ID | Representative PDB ID (chain) * in DB_{free} | Representative PDB ID (chain) * in DB_{bound} | Molecular name** | Used for DB_{free}**(b) *** |
|------------|---------------------------------------------|---------------------------------------------|-----------------|---------------------------|
| 97         | 2v1x(A)                                     | 2wyy(A)                                     | ATP-dependent DNA helicase Q1 | ✓             |
| 98         | 2ve8(A)                                     | 2ve9(A)                                     | DNA translocase FtsK | ✓             |
| 99         | 2vke(A)                                     | 1qpi(A)                                     | Tetracycline repressor | ✓             |
| 100        | 2wcw(A)                                     | 2wiw(B)                                     | Hjc              | ✓             |
| 101        | 2zwu(B)                                     | 3dnv(B)                                     | HTH-type transcriptional regulator HipB | ✓             |
| 102        | 2x6v(A)                                     | 3fde(A)                                     | E3 ubiquitin-protein ligase UHRF1 | ✓             |
| 103        | 2yve(A)                                     | 2yvh(A)                                     | Transcriptional regulator | ✓             |
| 104        | 2zd1(A)                                     | 3kk1(A)                                     | Reverse transcriptase p66 subunit | ✓             |
| 105        | 2zkg(A)                                     | 3fde(A)                                     | E3 ubiquitin-protein ligase UHRF1 | ✓             |
| 106        | 2zmz(A)                                     | 2el(A)                                      | Putative HTH-type transcriptional regulator PH1519 | ✓             |
| 107        | 3a45(A)                                     | 3a46(A)                                     | Formamidopyrimidine-DNA glycosylase | ✓             |
| 108        | 3bqz(A)                                     | 3fde(A)                                     | E3 ubiquitin-protein ligase UHRF1 | ✓             |
| 109        | 3bqz(A)                                     | 3fde(A)                                     | E3 ubiquitin-protein ligase UHRF1 | ✓             |
| 110        | 3bvs(A)                                     | 3jxy(A)                                     | Alkylpurine DNA glycosylase AlkD | ✓             |
| 111        | 3d6a(A)                                     | 3jxy(A)                                     | Alkylpurine DNA glycosylase AlkD | ✓             |
| 112        | 3d06(A)                                     | 3jxy(A)                                     | Alkylpurine DNA glycosylase AlkD | ✓             |
| 113        | 3d1g(A)                                     | 3jxy(A)                                     | Alkylpurine DNA glycosylase AlkD | ✓             |
| 114        | 3e5u(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |
| 115        | 3e1c(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |
| 116        | 3ebi(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |
| 117        | 3ebi(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |
| 118        | 3ebi(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |
| 119        | 3ebi(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |
| 120        | 3ebi(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |
| 121        | 3ebi(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |
| 122        | 3ebi(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |
| 123        | 3ebi(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |
| 124        | 3ebi(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |
| 125        | 3ebi(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |
| 126        | 3ebi(A)                                     | 3e6c(C)                                     | Cyclic nucleotide-binding protein | ✓             |

* A member which is reported with the highest resolution is shown as a representative. Entire lists of PDB ID (chain ID) of the members were shown in Table S2.

** The molecular names were extracted from the PDB headers of DNA-bound forms. If the molecular names in the headers do not describe the molecule (e.g. Putative protein), the molecular names were extracted from the literatures.

*** ✓ indicates that the clusters were used for DB_{free}**(b) ***

doi:10.1371/journal.pone.0056080.t001

Alphabet Propensities for the Conformationally Changed Fragments and for Disorder-to-order Conformationally Changed Fragments and Order-to-order Conformationally Changed Fragments

In a similar way as residue propensity, calculated are alphabet propensities in the DNA interfaces (\(R_{\text{bound/free(interface)}}\) and non-interfaces (\(R_{\text{bound/free(non-interface)}}\)) to characterize alphabets induced by conformational changes upon DNA binding, alphabet propensities for the DNA interfaces (\(R_{\text{bound/order~order~interface)}}\) and non-interfaces (\(R_{\text{bound/order~order~non-interface)}}\)) to characterize the order-to-order conformational changes, and alphabet propensities of the fragments that undergo a disorder-to-order conformational change upon DNA binding for the DNA interfaces (\(R_{\text{disorder/order,alp~i)}}\) and non-interfaces (\(R_{\text{disorder/order,non~interface)}}\)). The details for these calculations are provided as supplementary information (Text S1).

Statistical Reliability

Because the number of protein structures used in this study was limited, the statistical reliability of the calculated values was evaluated. The BCa bootstrap procedure [35] was used to estimate the confidence intervals for frequencies calculations on which propensities were calculated. We constructed 10,000 bootstrap datasets by resampling DB_{bound} and DB_{free}. In this test, the reliability standard was set as 85% of a two-sided confidence interval from the average value.

Results and Discussion

Dataset Preparation and Structural Alphabet Assignment

One hundred and twenty-six representative pairs of clusters in the DNA-free (DB_{free}) and DNA-bound (DB_{bound}) forms were...
obtained with a sequence similarity of less than 30%. The representative clusters were a set of subclusters with the largest members within each cluster (Table 1 and Table S1). The proteins of the 126 clusters had dsDNA binding domains that belonged to different structural classes according to the SCOP classification (version 1.75) [36]: 43 all alpha proteins, 12 all beta proteins, 30 alpha and beta proteins (α/β), 22 alpha and beta proteins (α+β), 11 multi-domain proteins (α and β), 1 small protein, and 1 coiled coil protein. The remaining 32 proteins were not classified in the SCOP database.

The obtained structures were divided into overlapped 4-residue-long fragments and assigned the relevant structural alphabets, and then, the changes in the alphabets in the DNA interfaces and non-interfaces were analyzed. The datasets contained 4963 fragments for the DNA interfaces and 20826 for the non-interfaces. If longer fragments are used, more fragments will be required to express the conformation within the similar range of errors to the 4-residue-long fragments; however, due to the limit of available data, 4-residue-long fragments were used to obtain statistically significant results.

Fragments in DNA Interfaces Tend to have more Intrinsic Variations in their Conformations than those in DNA Non-interfaces

The frequencies of conformational change upon DNA binding for the DBfree and DBbound datasets were calculated. A conformational change was considered to be a change in the alphabet between the DNA-free and DNA-bound forms. The frequencies of conformational changes in the DNA interfaces ($f_{\text{interface}}^{\text{chg}}$) and non-interfaces ($f_{\text{non-interface}}^{\text{chg}}$) were 23.1% and 14.7%, respectively (Fig. 3(a)). This result indicates that compared with non-interfaces, DNA interfaces tend to undergo more conformational change upon DNA binding.

It was anticipated that the fragments in DNA interfaces might have a more intrinsic propensity to change conformation in order to adjust to the DNA structure. To examine this assumption, the conformational variations for the fragments in the DNA-free forms were calculated using the dataset DBfree($\geq 2$), in which each cluster has at least two members in the DNA-free form. The frequencies of conformational variation in the DNA interfaces ($f_{\text{interface}}^{\text{var}}$) and non-interfaces ($f_{\text{non-interface}}^{\text{var}}$) are shown in Fig. 3(b). For the DNA-
No Specific Alphabets are Responsible for the Conformational Changes in DNA Interfaces

Next, the differences in the frequency of conformational change for the different alphabets were analyzed to reveal which local structures were affected most often. The alphabet propensities of conformational change in the DNA interfaces to that in the non-interfaces ($R_{\text{interface}}$) is one of the 11 alphabets) are shown in Fig. 4(a) and those for conformational variation ($R_{\text{var}}$) are shown in Fig. 4(b). In Fig. 4(a), a positive value indicates that frequency that alphabet has the conformational variation in DNA interfaces is higher than that in non-interfaces. In Fig. 4(b), a positive value indicates that frequency that alphabet has the conformational variation in DNA interfaces is higher than that in non-interfaces. As expected, conformational changes occurred more frequently in the DNA interfaces (positive values in Fig. 4(a)) for all the alphabets, and the alphabets A, D, E, F, G and Y of the 11 alphabets were significantly high. These 6 alphabets are likely to appear more often; however, the errors for B, C, H, I and J are too large to conclude that they significantly appear in the interfaces. The frequencies of the conformational variation in the DNA interfaces were also significantly higher than those for the non-interfaces except for G (Fig. 4(b)) though it is difficult to say which alphabet appears most often owing to the large errors.

Conformationally Changed Fragments in the DNA Interfaces have Amino Acids Suitable for Producing Flexibility and Binding to DNA

To reveal whether specific amino acids in the DNA interfaces affect the conformational change upon DNA binding, two pairs of residue propensities were calculated. First, the amino acid propensity in conformationally changed fragments to that in conformationally unchanged fragments ($R_{\text{chg} \to \text{unchg}}$) was determined for the DNA interfaces and non-interfaces (Fig. 5(a)). A positive value indicates that frequency that amino acid is observed in conformationally changed fragments is higher than that in unchanged fragments and zero indicates that both frequencies are equal. For example, 1.0 means that frequency of the amino acid in conformationally changed fragments is 2.7 times higher than that in unchanged fragments. Second, the amino acid propensity in the DNA interfaces to that in the non-interfaces was calculated for the conformationally changed and unchanged fragments (Fig. 5(b)).

Figure 5(a) shows that Asn, Gly, Pro, Ser, Asp, Thr and Lys have a positive value, indicating that they are favored amino acids in conformationally changed fragments located in non-interfaces. In contrast, the disfavored amino acids in the conformationally changed fragments of the non-interfaces were Trp, Cys, Ala, Arg, Leu, Met, Phe, Ile, and Val. These results clearly show that hydrophilic residues, Gly, and Pro are located on more flexible fragments in the non-interfaces. This trend was also found in disordered regions [11,12,31]. However, there were no significant differences between the propensity in the DNA interfaces and non-interfaces (filled and open bars in the figure), indicating that the conformation change in fragments depends basically on the amino acid types constituting the fragments and not on the positions.

The propensities of the amino acid frequency in the DNA interfaces against that in the non-interfaces for conformationally changed fragments ($R_{\text{chg} \to \text{unchg}}$) and conformationally changed fragments ($R_{\text{chg} \to \text{unchg}}$) are shown in Fig. 4(b). In Fig. 4(a), a positive value indicates that frequency that amino acid is observed in conformationally changed fragments is higher than that in unchanged fragments.
unchanged fragments \( \{R_{\text{interface}/\text{non-interface}}\} \) are shown in Fig. 5(b). A positive value indicates that frequency of amino acid in DNA interfaces is higher than that in non-interfaces and zero indicates that both frequencies are equal. The amino acids that favored to interact with DNA in the conformationally unchanged fragments are Arg, Thr, Gly, Ser, His, and Asn. On the other hand, Glu, Asp, Leu, Phe, Val, Ala, and Ile were disfavored in those fragments. The importance of basic and hydrophilic residues in the DNA interfaces has been emphasized in several previous reports [38,39,40]. Gly was also reported to be favored in protein–DNA interfaces but not in protein–protein interfaces [41]. Again, no significant differences could be detected in the propensity of the conformationally changed and unchanged fragments. These findings in amino acid propensities indicate that the amino acid preference depends solely upon the location of a fragment, that is, upon whether it is in a DNA interface or not, and whether a conformational change occurs upon DNA binding depends on the type of amino acids that constitute a fragment.

Three Specific Alphabets Appear more Often in Conformationally Changed Fragments Located in DNA Interfaces

Next, the alphabets that specifically increased upon DNA binding were evaluated to determine if they were different between the DNA interfaces and non-interfaces. To this end, the alphabet propensities \( R_{\text{bound}} = R_{\text{free}} \) for the DNA interfaces and \( R_{\text{bound}} = R_{\text{free}} \) for the non-interfaces were calculated and are shown in Fig. 6(a). Here, a positive value indicates that frequency that alphabet is observed in DNA-bound forms is higher than that in free forms. Error bar indicates the 85% bootstrap confidence interval.

doi:10.1371/journal.pone.0056080.g006

Figure 6. Propensities of the structural alphabets with various conditions. (a) Propensities of the structural alphabets that undergo a conformational change upon DNA binding. (b) Propensities of the structural alphabets that undergo an order-to-order conformational change upon DNA binding. (c) Frequencies of the structural alphabets that undergo a disorder-to-order conformational change (Y to one of A-J) upon DNA binding. In (a) through (d), fragments located within DNA interfaces and those within non-interfaces are shown by filled and open bar, respectively. A positive value of the propensities indicates that frequency that alphabet is observed in DNA-bound forms is higher than that in free forms. Error bar indicates the 85% bootstrap confidence interval.
and C are considered to be the top two preferred alphabets in disorder-order conformational changes in the DNA interfaces often occurred upon DNA binding. In particular, the conformations B, C, and H significantly increased compared with those in the non-interfaces. In contrast to the DNA interfaces, in the non-interfaces, loop-like conformations, the values of A (log propensity = 0.4), F, and G (0.2) were positive, and these conformations were induced upon DNA binding. Helix-like conformation I and extended conformation E (0.1) also slightly increased in the DNA-bound forms, whereas Y (~0.9) was disfavored. These results indicate that disordered fragments, even in non-interfaces, tend to be ordered when they bind to DNA.

Next, the reasons why the three above-mentioned conformations increased in the DNA interfaces were considered. Initially, it was recognized that conformational changes from Y to A–J significantly increased in the DNA interfaces compared with the non-interfaces. Therefore, it was expected that the distribution of the alphabets in the DNA interfaces would be more noticeably affected by disorder-to-order conformational changes. Figure 6(b) shows the propensities of the structural alphabets that undergo an order-to-order (that is, an A–J conformation to an A–J conformation) change upon DNA binding. Regardless of whether they were in a DNA interface, the values for all alphabets were nearly zero, indicating that there is no alphabet preference except for H. H increased in the DNA interfaces, but not in the non-interfaces in these order-to-order conformational changes. Thus, the protein–DNA complex structures with H were examined and the loop-like conformation (H conformation) was identified that stabilizes the protein–DNA interactions in various ways. However, owing to the limited number of data, no common features could be identified that explain why H increases upon DNA binding.

Next, for the fragments that underwent a disorder-to-order conformational change, the frequencies of the alphabets in the DNA interfaces and non-interfaces (Fig. 6(c)) and the alphabet propensities (Fig. 6(d)) were calculated. Neither the frequencies nor the propensities of the alphabets were significantly different between the DNA interfaces and the non-interfaces. This result indicates that the structures induced from disorder-to-order conformational changes in DNA interfaces are similar to those in non-interfaces. These results suggest that changes to B or C from Y in DNA interfaces occur more frequently. Consequently, B and C are considered to be the top two preferred alphabets in disorder-order conformational changes. The reasons why the B and C conformations are favored remains a subject for future investigation.

Conclusion

In this study, conformational changes in 4-residue fragments between DNA-free and DNA-bound forms were analyzed using structural alphabets, which enabled the precise description of the variety of local protein conformations. The results revealed the importance of the intrinsic conformational flexibility upon DNA binding: (1) intrinsic conformational variations in DNA interfaces are more frequent than those in non-interfaces and (2) conformationally changed fragments in DNA interfaces favor the disorder-promoting amino acids. In addition, it was found that three specific alphabets appeared in the DNA interfaces; however, the roles of the conformations in DNA binding are various. These findings may contribute to the more accurate prediction of the DNA binding sites of proteins and the potential conformational changes in the complex form.

Supporting Information

Table S1 The coordinates of the fragment library reported by Kolodny et al. (DOCX)

Table S2 PDB IDs and chains in DBfree, DBbound and DBfree>2- (XLXS)

Text S1 Supporting methods. (DOCX)

Acknowledgments

We thank Drs K. Ikebe, H. Ishida, T. Kai, N. Kanaeda, A. Matsumoto, S. Sakuraba, and Y. Yonetani for helpful advice and valuable discussions.

Author Contributions

Conceived and designed the experiments: TS HK. Performed the experiments: TS. Analyzed the data: TS HK. Contributed reagents/materials/analysis tools: TS HK. Wrote the paper: TS HK.

References

1. Nagadoi A, Morikawa S, Nakamura H, Enari M, Kobayashi K, et al. (1995) Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain. Structure 3: 1217–1224.
2. Brennan RG, Roderick SL, Takeda Y, Matthews BW (1996) Protein DNA Conformational-Changes in the Crystal-Structure of a Lambda-Cro-Operator Complex. Proc Natl Acad Sci U S A 87: 8165–8169.
3. Galburt EA, Chadsey MS, Jurica MS, Chevalier BS, Ezho D, et al. (2000) Conformational changes and cleavage by the homing endonuclease I-Ppol: A critical role for a leucine residue in the active site. J Mol Biol 300: 877–887.
4. Hegde RS, Wang AF, Kim SS, Schapira M (1998) Subunit rearrangement accompanies sequence-specific DNA binding by the bovine papillomavirus E2 protein. J Biol Chem 273: 797–808.
5. Tan S, Richmond TJ (1998) Crystal structure of the yeast MATalpha2/MCM1 DNA ternary complex. Nature 391: 660–666.
6. Hegde RS, Androphy EJ (1998) Crystal structure of the E2 DNA-binding domain from human papillomavirus type 16: implications for its DNA binding-site selection mechanism. J Mol Biol 284: 1479–1489.
7. Falconi M, Santalamazza A, Elieo T, De Prat-Gay G, Cicero DO, et al. (2007) Molecular dynamics of the DNA-binding domain of the papillomavirus E2 transcriptional regulator, uncover differential properties for DNA target accommodation. FEBS Lett 574: 2385–2393.
8. Falconi M, Oteri F, Elieo T, Cicero DO, Desideri A (2008) MD simulations of papillomavirus DNA-E2 protein complexes hints at a protein structural code for DNA deformation. Biophys J 95: 1104–1117.
9. Sauer RT, Smith DL, Johnson AD (1988) Flexibility of the yeast alpha 2 repressor enables it to occupy the ends of its operator, leaving the center free. Genes Dev 2: 807–816.
10. Voshod AK, Johnson AD (1995) A short, disordered protein region mediates interactions between the homodomain of the yeast alpha 2 protein and the MCM1 protein. Cell 72: 105–112.
11. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, et al. (2001) Intrinsically disordered protein. J Mol Graph Model 19: 26–59.
12. Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27: 527–533.
13. Liu J, Perumal NB, Oldfield C, Su EW, Uversky VN, et al. (2006) Intrinsically disordered in transcription factors. Biochemistry 45: 6873–6880.
14. Minezaki Y, Homma K, Kinjo AR, Nishikawa K (2006) Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. J Mol Biol 359: 1137–1149.
15. Dyson HJ (2012) Roles of intrinsically disordered protein-nucleic acid interactions. Mol Biosyst 8: 97–104.
16. Pauling L, Corey RB, Branson HR (1954) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci U S A 37: 205–211.
17. Dan A, Ofiran Y, Kliger Y (2009) Large-scale analysis of secondary structure changes in proteins suggests a role for disorder-to-order transitions in nucleotide binding proteins. Proteins 78: 236–248.
18. Mezei M (1998) Chameleon sequences in the PDB. Protein Eng 11: 411–414.
19. Offmann B, Tyagi M, De Bevern AG (2007) Local Protein Structures Current Bioinformatics 2: 163–202.
20. Le Q, Pollastru G, Krehl P (2009) Structural alphabets for protein structure classification: a comparison study. J Mol Biol 387: 431–450.
21. Deschavanne P, Tuffery P (2009) Enhanced protein fold recognition using a structural alphabet. Proteins 76: 129–137.
22. de Brevern AG, Etchebest C, Hazout S (2000) Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. Proteins 41: 271–287.
23. Etchebest C, Benros C, Hazout S, de Brevern AG (2005) A structural alphabet for local protein structures: improved prediction methods. Proteins 59: 810–827.
24. Friedberg I, Harder T, Kolodny R, Siben E, Li Z, et al. (2007) Using an alignment of fragment strings for comparing protein structures. Bioinformatics 23: e219–224.
25. Tsung CH, Huang JW, Yang JM (2007) Kappa-alpha plot derived structural alphabet and BLOSUM-like substitution matrix for rapid search of protein structure database. Genome Biol 8: R31.
26. Dudev M, Lim C. (2007) Discovering structural motifs using a structural alphabet: application to magnesium-binding sites. BMC Bioinformatics 8: 106.
27. Martin J, Regad L, Lecornet H, Camproux AC (2008) Structural deformation upon protein-protein interaction: a structural alphabet approach. BMC Struct Biol 8: 12.
28. Baussand J, Camproux AC (2011) Deciphering the shape and deformation of secondary structures through local conformation analysis. BMC Struct Biol 11: 9.
29. MacDonald JT, Makiniak K, Sadowski MI, Taylor WR. (2009) De novo backbone scaffolds for protein design. Proteins 78: 1311–1325.
30. Kolodny R, Koehl P, Guibas L, Levitt M (2002) Small libraries of protein fragments model native protein structures accurately. J Mol Biol 323: 297–307.
31. Linding R, Russell RB, Nedeva V, Gibson TJ (2003) GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 31: 3701–3708.
32. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235–242.
33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
34. Eisenhaber F, Argos P (1993) Improved strategy in analytic surface calculation for molecular systems: Handling of singularities and computational efficiency. Journal of Computational Chemistry 14: 1272–1280.
35. Efron B (1987) Better Bootstrap Confidence Intervals. Journal of the American Statistical Association 82: 171–175.
36. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247: 536–540.
37. Gunther S, Roether K, Frommel C (2006) Molecular flexibility in protein-DNA interactions. Biosystems 85: 126–136.
38. Jones S, van Heyningen P, Berman HM, Thornton JM (1999) Protein-DNA interactions: A structural analysis. J Mol Biol 287: 877–896.
39. Lejeune D, Delsaux N, Charloteaux B, Thomas A, Brasseur R (2003) Protein-nucleic acid recognition: statistical analysis of atomic interactions and influence of DNA structure. Proteins 51: 258–271.
40. Nadassy K, Wodak SJ, Janin J (1999) Structural features of protein-nucleic acid recognition sites. Biochemistry 38: 1999–2017.
41. Biswas S, Guharoy M, Chakrabarti P (2008) Structural segments and residue propensities in protein-RNA interfaces: comparison with protein-protein and protein-DNA complexes. Bioinformatics 2: 422–427.
42. Schrodinger L (2010) The PyMOL Molecular Graphics System, Version 1.3.