МОЛЕКУЛЯРНА БІОФІЗИКА

PROTEIN-DNA COMPLEXATION: CONTACT PROFILES IN DNA GROOVES

M.Yu. Zhitnikova[1], A.V. Shestopalova[1,2]

[1]*O. Ya. Usikov Institute for Radiophysics and Electronics, Natl. Acad. of Sci. of Ukraine,*
Acad. Proskury Street, 12, Kharkiv, 61085, Ukraine
[2]*V.N. Karazin Kharkiv National University, Svobody sq., 4, Kharkiv, 61022, Ukraine*
e-mail: allusio@gmail.com

Background: Investigation of the specific protein-DNA complexation mechanisms allows to establish general principles of molecular recognition, which must be taken into account while developing artificial nanostructures based on DNA, and to improve the prediction efficiency of the protein binding sites on DNA. One of the main characteristics of the protein-DNA complexes are the number and type of contacts in the binding sites of DNA and proteins. Conformational changes in the DNA double helix can cause changes in these characteristics.

Objectives: The purpose of our study is to establish the features of the interactions between nucleotides and amino acid residues in the binding sites of protein-DNA complexes and their dependence on the conformation of deoxyribose and the angle γ of the polynucleotide chain.

Materials and methods: At research of protein-DNA recognition process we have analyzed the contacts between amino acids and nucleotides of the 128 protein-DNA complexes from the structural databases. Conformational parameters of DNA backbone were calculated using the 3DNA/CompDNA program. The number of contacts was determined using a geometric criterion. Two protein and DNA atoms were considered to be in contact if the distance between their centers is less than 4.5 Å. Amino acid residues were arranged according to hydrophobicity scale as hydrophobic or nonpolar and polar.

Results: The analysis of contacts between polar and hydrophobic residues and nucleotides with different conformations of the sugar-phosphate backbone showed that nucleotides form more contacts with polar amino acids in both grooves than with hydrophobic ones regardless of nucleotide conformation. But the profile of such contacts differs in minor and major grooves and depends on the conformation of both deoxyribose and γ angle. The contact profiles are characterized by the sequence-specificity or the different propensity of nucleotides to form contacts with the residues in both grooves.

Conclusions: Our analysis have shown, that the amount and type of protein-nucleic contacts and their distribution in the grooves depend on the conformation of the sugar-phosphate backbone, the nucleotide sequence and the type of amino acids in the binding sites.

KEY WORDS: protein-DNA recognition; protein-DNA contacts; DNA structure; binding site prediction; structural variations.

КОМПЛЕКСОУТВОРЕННЯ БІЛКИ-ДНК: КОНТАКТНІ ПРОФІЛІ В ЖОЛОБКАХ ДНК
М.Ю. Житнікова[1], Г.В. Шестопалова[1,2]

[1]*Інститут радіофізики та електроніки ім. О.Я.Усикова НАН України,*
вул. Акад. Проскури, 12, Харків, 61085, Україна
[2]*Харківський національний університет імені В.Н. Каразіна, м. Свободи, 4, 61022, Харків, Україна*

Актуальність. Дослідження механізмів специфічного білково-нуклеїнового комплексоутворення дозволяє встановити загальні принципи молекулярного впізнавання, які необхідно враховувати при розробці штучних наноструктур на основі біополімерів, а також для підвищення ефективності прогнозування сайтів зв'язування білків і ДНК. Одними з основних характеристик білково-нуклеїнових комплексів є кількість і тип контактів в сайтах зв'язування ДНК і білків. Конформаційні перебудови подвійної спіралі ДНК можуть викликати зміни цих характеристик.

Мета роботи. Метою нашого дослідження було визначення особливостей взаємодії між нуклеотидами і амінокислотними залишками в сайтах зв'язування білково-нуклеїнових комплексів в залежності від конформації дезоксирибози і кута γ полінуклеотидного ланцюга.

Матеріали і методи. При дослідженні механізмів білково-нуклеїнового впізнавання ми встановили міжмолекулярні контакти для 128 білково-нуклеїнових комплексів із структурних баз даних. Конформаційні параметри цукрофосфатного остову ДНК розраховувалися за допомогою пакету програм 3DNA/CompDNA. Кількість білково-нуклеїнових контактів визначалася за

геометричним критерієм. Два атома білка і ДНК вважалися взаємодіючими, якщо відстань між їх центрами була менша 4,5 Å. Амінокислотні залишки позначалися за шкалою гідрофобності як гідрофобні або неполярні та полярні.

Результати. Аналіз розподілу білково-нуклеїнових контактів показав, що нуклеотиди, незалежно від конформації, утворюють в обох жолобках більше контактів з полярними амінокислотними залишками, ніж з гідрофобними. Але профіль таких контактів різниться в малому і великому жолобках і залежить як від конформації дезоксирибози, так і кута γ. Контактні профілі також характеризуються сиквенс-специфічністю або різною схильністю нуклеотидів до взаємодій з амінокислотними залишками в обох жолобках.

Висновки. Кількість і тип білково-нуклеїнових контактів та їх розподіл по жолобках залежать від конформації цукрофосфатного остову, нуклеотидної послідовності і типу амінокислот в сайтах зв'язування.

КЛЮЧОВІ СЛОВА: білково-нуклеїнове впізнавання; білково-нуклеїнові контакти; структура ДНК; передбачення сайтів зв'язування; структурні варіації.

КОМПЛЕКСООБРАЗОВАНИЕ БЕЛКИ-ДНК: КОНТАКТНЫЕ ПРОФИЛИ В ЖЕЛОБКАХ ДНК
М.Ю. Житникова[1], А.В. Шестопалова[1,2]
*[1]Институт радиофизики и электроники им. А.Я.Усикова НАН Украины,
ул. Акад.Проскуры, 12, Харьков, 61085, Украина
[2]Харьковский национальный университет имени В.Н. Каразина, пл. Свободы, 4, 61022, Харьков, Украина*

Актуальность. Исследование механизмов специфического белково-нуклеинового комплексообразования позволяет установить общие принципы молекулярного узнавания, которые необходимо учитывать при разработке искусственных наноструктур на основе биополимеров, а также для повышения эффективности прогнозирования сайтов связывания ДНК и белков. Одними из основных характеристик белково-нуклеиновых комплексов являются количество и тип контактов в сайтах связывания ДНК и белков. Конформационные перестройки двойной спирали ДНК могут вызывать изменения этих характеристик.

Цель работы. Целью нашего исследования было определение особенностей взаимодействия между нуклеотидами и аминокислотными остатками в сайтах связывания белково-нуклеиновых комплексов в зависимости от конформации дезоксирибозы и угла γ полинуклеотидной цепи.

Материалы и методы. При исследовании механизмов белково-нуклеинового узнавания мы установили межмолекулярные контакты для 128 белково-нуклеиновых комплексов из структурных баз данных. Конформационные параметры сахаро-фосфатного остова рассчитывались с помощью пакета программ 3DNA/CompDNA. Количество контактов определялось с помощью геометрического критерия. Два атома белка и ДНК считались взаимодействующими, если расстояние между их центрами меньше 4,5 Å. Аминокислотные остатки распределялись по шкале гидрофобности как гидрофобные или неполярные и полярные.

Результаты. Анализ распределения белково-нуклеиновых контактов показал, что нуклеотиды, независимо от конформации, образуют в обоих желобках больше контактов с полярными аминокислотными остатками, чем с гидрофобными. Но профиль таких контактов различается в малом и большом желобках и зависит как от конформации дезоксирибозы, так и угла γ. Контактные профили также характеризуются сиквенс-специфичностью или различной склонностью нуклеотидов к взаимодействиям с аминокислотными остатками в обоих желобках.

Выводы. Количество и тип белково-нуклеиновых контактов и их распределение по желобкам зависят от конформации сахарофосфатного остова, нуклеотидной последовательности и типа аминокислот в сайтах связывания.

КЛЮЧЕВЫЕ СЛОВА: белково-нуклеиновое узнавание; белково-нуклеиновые контакты; структура ДНК; предсказание сайтов связывании; структурное многообразие.

Molecular recognition of DNA by proteins is a fundamental problem in molecular biology and drug design. Investigation of the mechanisms of specific protein-DNA complexation allows us to establish general principles of the molecular recognition, which must be taken into account while developing artificial nanostructures based on DNA, and, also, to improve the efficiency of the prediction of protein-DNA binding sites. The prediction of DNA binding sequences for proteins is an unsolved problem, but significant progress has occurred in the last years [1]. To improve the reliability of DNA-binding sites prediction, some problems must be solved. Primarily

structural and physicochemical complementarity between DNA sequences and proteins or between the DNA nucleotides and interacting amino acid residues are the principal factors of specificity. That is the binding site information is depending on both protein and DNA structure [2]. For most of protein-DNA complexes the location and sequences of binding sites are not defined. Their experimental determination requires expensive methods. But if we have the information about the spatial structures of the complexes, it is possible for us to predict *in silico* protein-DNA binding sites [3-5].

One of the characteristics of protein-DNA complexation is the number and type of the contacts in protein-DNA binding sites. The detailed analysis of protein-DNA interactions are presented in Refs. [6,7]. Authors have determined all protein-nucleic contacts and separated them according to the types of interactions and possible places of their formation in complexes. They have found that the most common types of protein-nucleic interactions are hydrogen bonds and van der Waals contacts. As for the primary binding sites it turned out that more than two-thirds of the protein contacts are formed with the DNA sugar–phosphate backbone.

Simultaneously conformational changes in DNA can lead to changes in the protein-DNA contact profiles, which are formed, as a rule, in the grooves of the double helix [8, 9]. Since, the DNA shape can implicate to protein-DNA binding specificity [9], we also have to look for any peculiar properties of DNA shape.

It should be noted that in protein-DNA binding sites is observed an enriched occurrence of the A-like conformers and mixed A/B conformers [10]. The tendency of A-like nucleotides appearance in surface of protein-DNA complexes is accompanied by a transition of the deoxyribose from the C2'-endo (B-like) to C3'-endo (A-like) sugar pucker. This effect is known as the 'sugar switching' and facilitating by hydrophobic interactions in the minor groove [8, 11]. Unusual DNA conformers are often concentrated in the DNA binding sites of protein-DNA complexes. Obviously the propensity of DNA to adopt unusual conformers is essential for their recognition by proteins. The occurrence of these non-canonical forms of the DNA double helix can also explain the role of such DNA structures for evolution [12].

Typical re-arrangements of the DNA sugar-phosphate backbone in the complexes with protein involve B→A sugar switching and rotations around the four torsion angles (α, γ, ε, ζ) with following possible conformations: *gauche+* (*g+*), *trans* (*t*), *gauche-* (*g-*) [8,13,14]. Earlier [15] we focused on the question of how the transition of the γ angle (O5'-C5'-C4'-C3') in the DNA sugar-phosphate backbone from classical *g+* to alternative *t* and *g-* conformations affects the local DNA interface accessible for interaction with proteins in the minor and major grooves and evaluate the sequence dependence of such transitions?

The purpose of this study is to establish the features of the interactions between nucleotides and amino acid residues in protein-DNA binding sites depending on the conformation of deoxyribose and the angle γ of the polynucleotide chain. The results are based on the statistical analysis of crystallographic structures of protein-DNA complexes from the NDB database [16].

MATERIALS AND METHODS

Data set of crystallographic structures

The data-set of the 128 protein-DNA complexes was generated for comprehensive analysis. For the structure selection we use crystallographic coordinates of the protein-DNA complexes extracted from the Nucleic acids Data Bank [16].

The resolution of structures solved by the X-ray crystallography is better than 1.9 Å. This resolution has been previously identified as the one that ensures accurate determination of the

sugar puckers and backbone torsion angles as well as statistical analysis of their distribution [13]. All chosen complexes were defined as any structure containing one or more protein chains and at least one double-stranded DNA with more than 4 nucleotides in each chain. From this set we excluded modified and terminal unpaired nucleotides. In order to avoid the end effects, the terminal base pairs on both ends of each DNA structure are excluded from the analysis. Moreover, the data-set is non-redundant, i.e. only one structure is selected from the group of complexes containing an identical protein or its mutants. The preference is given to the complexes with wild-type proteins and better resolution.

Calculation of the DNA structural parameters

The values of the dihedral angle γ of the DNA backbone (O5′–C5′–C4′–C3′) and the angle of pseudorotation (P) of the sugar ring for each nucleotide are calculated by means of the 3DNA/CompDNA analyzer [17].

The angle γ values are classified as *gauche+ (g+)*, *trans (t)*, and *gauche- (g-)* according to the classical threefold staggered pattern of the dihedrals: *g+* (60° ± 30°), *t* (180° ± 30°), *g-* (300° ± 30°). The sugar pucker is sorted into the categories of «south» B-like (120° < P < 220°) and «north» A-like sugars (-60° < P < 60°). The nucleotides with stereochemically unfavorable P and γ values outside of these domains are not included into analysis.

Revealing the atom–atom contacts in the protein–DNA binding sites

An amino acid residue is considered as forming contacts if the distance between the atoms of the residues and the nucleotide atoms is less than a cutoff radius. Several studies have used different methods of definitions of protein-DNA contacts in the binding sites [1, 18]. Two atoms of the residue and nucleotide are considered to form contact if the distance between their centers is less than 4.5Å [7, 8] and afterward authors in Ref. [19] declared that this cutoff distance gave the best determination of binding and nonbinding residues to predict binding sites.

We also use the cutoff value of 4.5 Å in our investigation. In this way, we find atoms interacting according to all different mechanisms (hydrogen bonds, electrostatic interactions, and van der Waals interactions). We identify the contacting pairs of atoms in the protein–DNA binding sites by means of the modified software package proposed in [8] that distinguishes contacts in the major and minor grooves. The atoms of bases are also divided into groups exposed in the major (N4, N6, N7, O4, O6, C5, C6, C7, C8, and C4 in Thymine and Cytosine) and minor grooves (N2, N3, O2, C2, and C4 in Adenine and Guanine). The nucleotides which contain atoms contacting with protein are classified as interacting nucleotides, with further subdivision into those interacting in the minor and major grooves.

Twenty amino acids have different hydrophobic and polar properties [20,21]. In our analysis all amino acid residues are separated as hydrophobic or nonpolar IFVLMAGC and polar WYTPSHENQDKR residues (denote by one-letter codes) according to the hydrophobicity scale [22].

RESULTS AND DISCUSSION

Distribution of the protein-DNA contacts

In previous investigation [15] we have shown that the ability of the nucleotides to participate in contacts with protein residues for the both grooves is connected with conformations of their sugar moiety and torsion angle γ. Now more detail analysis presented in this investigation allowed us to make some clarifications. The protein–DNA interactions were calculated for 128 protein-DNA complex structures. These complexes include 1765 nucleotides: 230 A-like nucleotides

and 1535 B-like ones. The nucleotides in the classical conformation (1352 nucleotides with the B-conformation of deoxyribose and *g+* conformation of the γ angle) are the most frequently found, while the other conformations are much less frequent. For A-like nucleotides there are: 152 *g+* classical and 71 *t* alternative nucleotides, for B-like: 1352 *g+*, 53*g-* and 38 *t*. However, as previously reported [8,9,13-15], such alternative conformations can play an important role in the process of indirect proteins recognition of their binding site on DNA sequence. The more detailed scheme of the structure selection and base composition can be found in Ref. [15].

In total we have analyze the 26293 amino acid–DNA contacts. The numbers of contacts between amino acid and DNA in the both grooves are counted separately for bases and sugar-phosphate backbone. Then, all nucleotides were divided according to their deoxyribose conformation (the A-like or the B-like) and conformation of γ angle (classical *g+* and alternative *t* or *g-*).

At first we consider the distribution of the protein-DNA interactions between the grooves. According to our data (Table 1) A-like nucleotides more often interact with proteins in a minor groove (3853 (2351 *g+*) and 2793 (1682 *g+*) contacts in minor and major groove, respectively), while B-like nucleotides more often form contacts with protein residues in the major groove (8446 (7918 *g+*) and 11935 (10368 *g+*) contacts in minor and major groove respectively) regardless of the γ angle conformation. But it should be noted that alternative A-like nucleotides form more contacts in the grooves (1503 and 1111 contacts in minor and major groove, respectively) than alternative B-like nucleotides with both *t* and *g-* angle γ conformations (528 and 967 contacts in minor and major groove, respectively).

Table 1

Number of amino acid residue contacts with bases and sugar-phosphate backbone for different nucleotides in the both grooves

	Minor groove					Major groove				
	A-like		B-like			A-like		B-like		
	g+	*t*	*g+*	*t*	*g-*	*g+*	*t*	*g+*	*t*	*g-*
	Base contacts									
A	106	20	244	0	0	105	48	774	21	15
C	73	12	107	7	0	336	94	757	75	27
G	44	74	394	25	13	119	247	1342	37	52
T	72	24	127	0	12	151	88	1071	30	67
	Sugar-phosphate backbone contacts									
A	455	362	1858	9	28	186	96	1507	160	28
C	791	164	1557	157	25	401	140	1171	100	68
G	348	599	2196	86	42	131	225	1958	75	70
T	462	248	1435	53	71	253	173	1653	64	78

Hence the conformational transitions of sugar-phosphate backbone do not interfere with the formation of the protein-DNA contacts in both grooves even facilitating their realization in some particular situations (see below).

Then we examine the location of contacts in the grooves and determine that near 3/4 of the protein-DNA contacts are formed with the sugar-phosphate backbone (19483 contacts or 74%), and the rest one (6810 contacts or 26%) are formed with bases. At the same time the distribution of the contacts between the grooves are differed. In the minor groove only 10% of interactions are the contacts between residues and bases, and 90% of residues contacts are formed with sugar-phosphate backbone. In the major groove the number of contacts with bases increases up to 39%. These results are in agreement with the well-known idea that specific protein contacts with DNA

bases are realized in the major groove whereas nonspecific interactions with the sugar-phosphate backbone are occurred in the minor groove [2,6,8,].

But the percentage of contacts between protein and bases or backbone atoms in the grooves is different for nucleotides with various conformations of the sugar-phosphate backbone. In the minor groove, the transition to alternative *t* conformation of the angle γ increases the participation of backbone atoms in the contacts with proteins, especially for A-like nucleotides: from 88% for *g*+ state to 91.4% for *t* state. In the major groove, only the transition of B-like nucleotides to *t* γ angle conformation alters the participation of nucleotides in the contacts, increasing the number of backbone atom contacts (from 62 to 71%).

That is, transitions to alternative conformations of the angle γ can influence the features of protein-nucleic interactions primarily in the minor groove.

We also can notice some common patterns in the sequence-specificity of contacts for nucleotides with different sugar-phosphate backbone conformations.

In total the greatest number of contacts with amino acid residues forms nucleotides containing Guanine, while the other three types of nucleotide create practically the equal number of contacts. The decreasing order of contact propensity is: Guanine > Thymine > Cytosine ≥ Adenine. This result is explained by the fact that Guanine exposes the greatest number of potential hydrogen-bonding atoms on the base edges [6] and only Guanine can serve as a hydrogen bond donor in the minor groove [23].

Conformational angle γ transitions and sugar switching lead to a different participation of all four types of nucleotides in the protein-DNA contacts.

It should be noted, that B-like alternative nucleotides contained Guanine are involved in contacts between Guanine and amino acid residues in the minor groove while B-like alternative nucleotides containing Adenine, do not form any contacts between Adenine and amino acid residues. Interesting, that Thymine nucleotides with B-like deoxyribose and *g-* γ angle conformation more often take part in the protein-DNA contacts in comparison to other conformations. Obviously this conformation promotes the interaction of the Thymine bases with amino acids, which can be one of the ways of the indirect recognition mechanism realization.

For A-like nucleotides in both grooves the transition to alternative *t* angle γ conformation leads to even the greater increasing of the number of Guanine contacts and for their decreasing with other bases, especially with Cytosine. The similar tendency is observed for amino acid contacts with sugar-phosphate backbone of A-like nucleotides: the number of contacts is increased for nucleotides containing Guanine and decreased for the other types of nucleotides.

In the major groove B-like nucleotides in *t* γ angle state containing Adenine form more backbone contacts than other types of nucleotides with the similar sugar-phosphate conformation. At the same time Adenines create the fewer number of base contacts in major groove as well as Cytosines in minor groove (both with alternative γ angle conformations).

Thus the contact profiles are characterized by the sequence-specificity, that is, the different propensity of certain nucleotides to form contacts with amino acids in the grooves.

Contacts with polar and hydrophobic amino acid residues

The analysis of contacts between polar and hydrophobic amino acid residues and nucleotides with different conformations of the sugar-phosphate backbone shows that nucleotides form more contacts with polar amino acids in both grooves than with nonpolar residues regardless of the nucleotide conformation (Table 2, Total contacts).

But, the profiles of such contacts differ in minor and major grooves and depend on the conformation of both deoxyribose and the angle γ (Fig.1).

The nucleotides with classical *g*+ conformation of the γ angle are characterized by the similar contact profiles in both grooves independently on the deoxyribose conformation.

There are slightly changes in contacts profiles for A-like nucleotides with alternative *t* conformation of the γ angle (Table 2). The percent of contacts with polar amino acids in minor groove increases, and with hydrophobic amino acids decreases. In major groove, on the contrary, the percent of contacts with polar amino acids decreases, and with hydrophobic amino acids increases. These facts are in agreement with the change in the polar/nonpolar profile of the DNA surface, available in the major and minor grooves [15].

Table 2

Percent of contacts between polar/nonpolar amino acids and nucleotides with different γ angle conformations in the both grooves

			Total contacts		Bases		Sugar-phosphate backbone	
			Nonpolar amino acids	Polar amino acids	Nonpolar amino acids	Polar amino acids	Nonpolar amino acids	Polar amino acids
Major groove	A-like	*g+*	12.7	87.3	6.9	93.1	17.0	83.0
		t	14.0	86.0	13.2	86.8	14.6	85.4
	B-like	*g+*	15.8	84.2	11.1	88.9	18.7	81.3
		t	8.2	91.8	5.5	94.5	**9.3**	**90.7**
		g-	16.5	83.5	18.0	82.0	15.6	84.4
Minor groove	A-like	*g+*	25.4	74.6	43.1	56.9	22.8	77.2
		t	21.2	78.8	19.2	80.8	21.3	78.7
	B-like	*g+*	21.6	78.4	18.6	81.4	22.0	78.0
		t	20.2	79.8	21.9	78.1	20.0	80.0
		g-	29.4	70.5	**80.0**	**20.0**	21.7	78.0

More careful evaluation of polar/hydrophobic contact profile shows that for A-like nucleotides with alternative *t* conformation of the γ angle the percent of residue contacts is differ for bases and backbone atoms (Table 2). For B-like nucleotides with alternative *t* conformation of the γ angle in both grooves, the percent of contacts with polar amino acids increases, and with hydrophobic amino acids decreases, especially in major groove. But the transition of B-like nucleotides to the *g-* alternative conformation reduces the percent of contacts with polar amino acids and increases it with nonpolar ones in both grooves (Table 2, Total contacts). So we observe the significant changes in their contact profiles (Fig.1). The verification of the results suggests that we need to be careful interpreting the high propensities of hydrophobic amino acids contact with bases of nucleotides in this conformation (B-like sugar pucker, *g-* conformation) because there are only 25 total residue-base contacts in the whole dataset. Whatever, such minor groove contacts may play a role for proteins binding only in the minor groove (such as Architectural), which is often associated with a dramatic widening and extensive hydrophobic contacts [24] and contribute to specificity.

The analysis of contacts has shown, that the significant changes are observed for base contacts rather than for sugar-phosphate backbone ones and primarily for classical nucleotides independently of deoxyribose conformation (Table 2). For alternative nucleotides such tendency is less defined. Only one exception is observed in the major groove for B-like nucleotides with *t* conformation representing the significant changes in sugar-phosphate backbone contact profiles. The most dramatical changes observed for the base contact profiles in the minor groove for B-like nucleotides with *g-* angle γ conformation we have discussed above.

Conformational transitions of sugar-phosphate DNA backbone can affect the propensity for DNA-binding of amino acids (Fig. 1).

Our determination the propensity of amino acids to interact with nucleotides shows, that the most common the protein-DNA contacts form positive charged Arginine (R) and Lysine (K) residues (Fig. 1) and this data are consistent with the results of many studies [6, 7, 23].
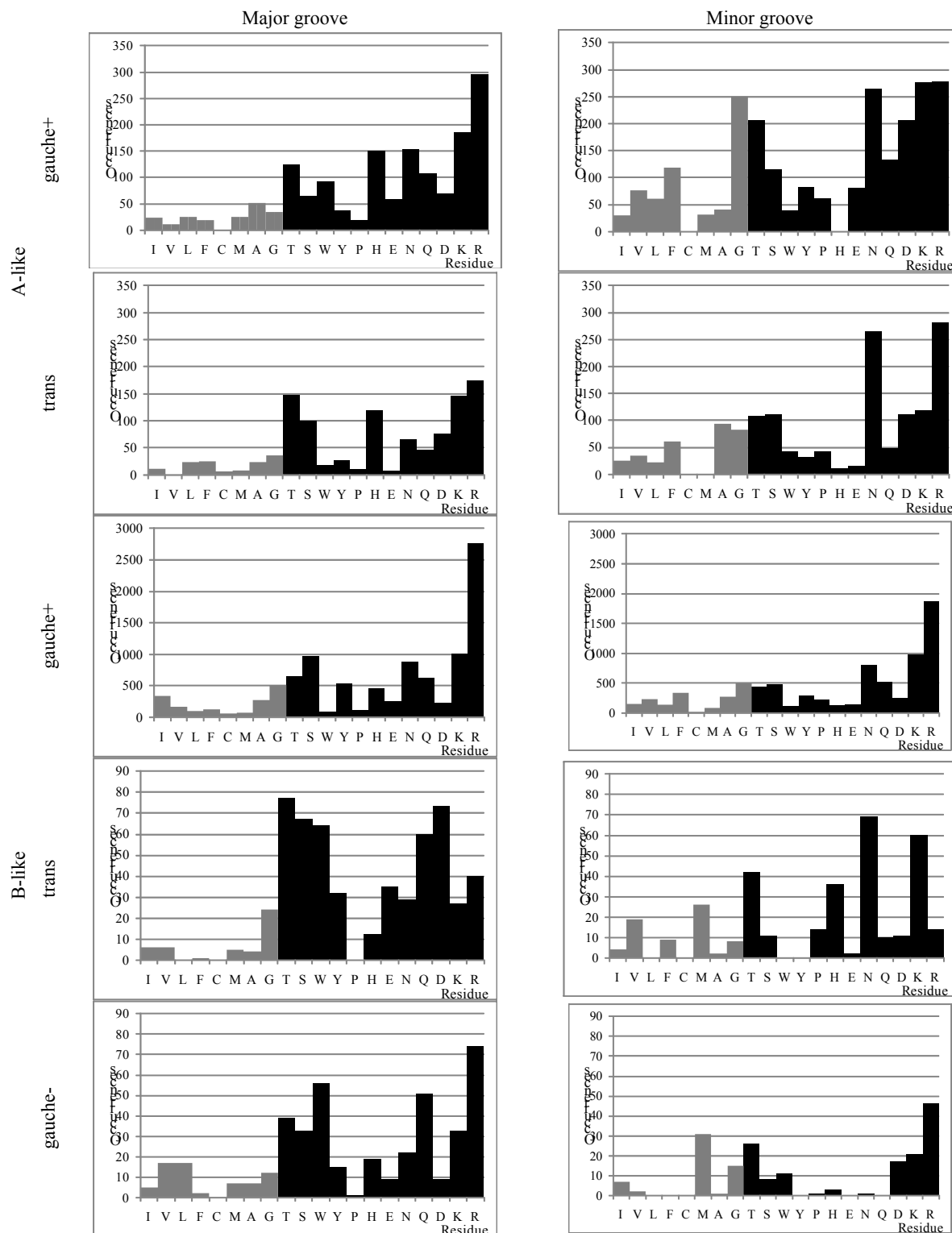

Fig. 1. Contact profiles in major and minor grooves or Number of the protein-DNA contacts between nonpolar (IVLFCMAG, grey) and polar (TSWYPHENQDKR, black) amino acids and nucleotides with different sugar-phosphate backbone conformations: A- or B-like deoxyribose; gauche+, trans, gauche- states of angle γ.

Tryptophan (W) has the largest propensity to form contacts with the DNA in major groove for B-like nucleotides in alternative γ angle conformations. Despite the fact that Tryptophan is referred as polar amino acid [22], the presence of a heterocyclic system in its side chain allows Tryptophan to participate in hydrophobic interactions. Therefore the observed effect can be explained by the increasing of nonpolar surface in the major groove upon the transitions of γ angle into alternative conformations [15] facilitating the formation of the specific hydrophobic contacts with nucleotides [25].

The number of Aspartate (D) contacts in DNA major groove is high for nucleotides with any sugar pucker and *t* γ angle conformations (Fig.1). Moreover, almost half of the contacts in minor groove for nucleotides with B-like sugar and *t* γ angle conformation are the Cytosine interactions (Table 1). Recently Corona et al. [23] have shown that negatively charged Aspartate is enriched in base interactions for highly specific DNA-binding proteins and predominately binds to Cytosine in the major groove through a single hydrogen bond or two consecutive Cytosines through bidentate hydrogen bonds. Obviously, transitions into *t* angle γ conformation contribute binding specificity through indirect readout mechanism.

The transition of B-like nucleotides into *g-* angle γ conformation reduces the number of Arginine (R) contacts in minor groove (Fig.1). This effect can be associated with sequence-specificity. Arginine is prefer to bind with Guanine [6], while we have shown, that in this conformation Thymine is more often takes part in the protein-DNA contacts (Table 1).

CONCLUSIONS

Prediction of protein binding sites on DNA will allow us to determine the functions of proteins and understand regulatory processes in molecular biological systems as well as to develop pharmaceutical drugs that can prevent the expression of target genes. For solving this problem some general principles concerning the frequency of formation of the specific amino acid-base pairs in binding sites should be formulated.

In this investigation we try to identify common features of DNA backbone rearrangements which can effect on the ability of nucleotides to participate in contacts with proteins. According to the results of our study one can assert the existence of certain preferences in the formation of the protein-DNA contacts, depending on the conformation of the DNA sugar-phosphate backbone.

The A-like nucleotides more often interact with proteins in the minor groove, while B-like nucleotides make more contacts in the major groove regardless of the γ angle conformation. At the same time the alternative A-like nucleotides form more contacts in the grooves than B-like nucleotides with both *t* and *g-* alternative angle γ conformations.

Near 3/4 of the protein-DNA contacts are formed with sugar-phosphate backbone. Such interactions predominate in the minor groove especially for A-like alternative nucleotides. In the major groove the number of contacts with bases is significantly higher, but the transition of B-like nucleotides to alternative *t* conformation of angle γ leads to increasing the interaction with backbone atoms.

The nucleotides form more contacts with polar amino acids in both grooves than with nonpolar residues regardless of the nucleotide conformation, but the profiles of such contacts differ in minor and major grooves and depend on the conformation of both deoxyribose and the angle γ.

In the minor groove the amount of interactions between alternative nucleotides and polar amino acids increase, and with hydrophobic amino acids decrease. In the major groove, on the contrary, alternative nucleotides form less number of contacts with polar amino acids than with hydrophobic ones. The polar/hydrophobic contact profile for alternative A-like nucleotides is differed for bases and backbone atoms.

We have shown that contact profiles are characterized by the sequence-specificity, that is, the different propensity of certain nucleotides to form contacts with amino acids in both grooves. In particular the transitions to alternative conformations of the angle γ can influence the specificity of the protein-DNA interactions primarily in the minor groove. But the correct determination of preferred amino acid − nucleotide pairs in the binding sites of protein-DNA complexes requires additional studies.

Statistical analysis of the presented protein-DNA structures allows us to make following concluding remark. It is well known that there are no simple rules in protein-DNA recognition [9], and the prediction of binding sites is extremely difficult, since proteins and DNA have a significant conformational variability. Most of the methods predicting protein binding sites on DNA are based on sequence preferences of residues and protein structure in binding site [1] and take into account only the DNA sequence (i.e., direct readout) [26-31]. We assume that information about conformational rearrangements of DNA (indirect or shape recognition) will allow to supplement the results of these methods and take into consideration not only protein features but also the structure of DNA. Such data can substantially improve the quality of binding sites prediction.

Acknowledgment

REFERENCES

1. Si J., Zhao R., Wu R. An Overview of the Prediction of Protein DNA-Binding Sites // Int. J. Mol. Sci. 2015. No 16. P. 5194-5215.
2. Mandel-Gutfreund Y., Margalit H. Quantitative parameters for amino acid–base interaction: implications for prediction of protein–DNA binding sites // Nucleic Acids Res. 1998. Vol. 26. No 10. P. 2306–2312.
3. Jones S., Shanahan H.P., Berman H.M., Thornton J.M. Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins // Nucleic Acids Res. 2003. No 31. P. 7189–7198.
4. Shanahan H.P., Garcia M.A., Jones S., Thornton J.M. Identifying DNA-binding proteins using structural motifs and the electrostatic potential // Nucleic Acids Res. 2004. No 32. P. 4732–4741.
5. Ofran Y., Mysore V., Rost B. Prediction of DNA-binding residues from sequence // Bioinformatics. 2007. Vol. 23. No 13. P. i347–i353.
6. Luscombe N.M., Laskowski R.A., Thornton J.M. Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level // Nucleic Acids Research. 2001. Vol. 29. No 13. P. 2860–2874.
7. Lejeune D., Delsaux N., Charloteaux B., Thomas A., Brasseur R. Protein–Nucleic Acid Recognition: Statistical Analysis of Atomic Interactions and Influence of DNA Structure // Proteins. 2005. Vol. 61. No 2. P. 258–271.
8. Tolstorukov M.Y., Jernigan R.L., Zhurkin V.B. Protein-DNA hydrophobic recognition in the minor groove is facilitated by sugar switching // J. Mol. Biol. 2004. No 337. P. 65–76.
9. Rohs R., Jin X., West S.M., Joshi R., Honig B., Mann R.S. Origins of specificity in protein-DNA recognition // Annual review of biochemistry. 2010. No 79. P. 233–26.
10. Schneider B., Cerny J., Svozil D., Cech P., Gelly J.-C., de Brevern A. G. Bioinformatic analysis of the protein/DNA interface // Nucleic Acids Research. 2014. Vol. 42. No 5. P. 3381–3394.
11. Locasale J.W., Napoli A.A., Chen S., Berman H.M., Lawson,C.L. Signatures of protein-DNA recognition in free DNA binding sites // J. Mol. Biol. 2009. No 386. P. 1054–1065.
12. Zhao J., Bacolla A., Wang G., Vasquez K.M. Non-B DNA structure-induced genetic instability and evolution // Cell. Mol. Life Sci. 2010. No 67. P. 43–62.
13. Schneider B., Neidle S., Berman H. M. Conformations of the sugar–phosphate backbone in helical DNA crystal structures // Biopolymers. 1997. Vol. 42. No 1. P. 113–124.
14. Varnai P., Djuranovic D., Lavery R., Hartmann B. α-γ transitions in the B-DNA backbone // Nucl. Acids Res. 2002. Vol. 30. No 24. P. 5398–5406.
15. Zhitnikova M.Yu., Boryskina O.P., Shestopalova A.V. Sequence-specific transitions of the torsion angle gamma change the polar-hydrophobic profile of the DNA grooves: implication for indirect protein-DNA recognition // J. Biomol. Struct. Dyn. 2014. Vol. 32. No 10. P. 1670-1685.
16. Berman H. M., Olson W. K., Beveridge D. L., Gelbin A., Demeny T., Hsieh S. H., Srinivasan A.R., Schneider B. The nucleic acid database: A comprehensive relational database of three-dimensional structures of nucleic acids // Biophysical Journal. 1992. No 63. P. 751–759.

17. Lu X. J., Olson W. K. 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures // Nucleic Acids Research. 2003. No 31. P. 5108–5121.

18. Ahmad S., Gromiha M.M., Sarai A. Analysis and prediction of DNA-binding proteins and their binding residues based on composition, sequence and structural information // Bioinformatics. 2004. No 20. P. 477–486.

19. Kuznetsov I.B., Gou Z., Li R., Hwang S. Using evolutionary and structural information to predict DNA-binding sites on DNA-binding proteins // Proteins. 2006. No 64. P. 19–27.

20. Bartlett G.J., Porter C.T., Borkakoti N., Thornton J.M. Analysis of catalytic residues in enzyme active sites // J. Mol. Biol. 2002. No 324. P. 105–121.

21. Faucher J., Pliska V. Hydrophobic parameters pi of amino acid side chains from the partitioning of N-acetyl-amino-acid amides // Eur. J. Med. Chem. 1983. No 18. P. 369–375.

22. Kyte J., Doolittle R.F. A simple method for displaying the hydropathic character of a protein // J. Mol. Biol. 1982. No 157. P. 105–132.

23. Corona R. I., Guo J.T. Statistical analysis of structural determinants for protein-DNA binding specificity // Proteins. 2016. Vol. 84. No 8. P. 1147–1161.

24. Bewley C.A., Gronenborn A.M., Clore G.M. Minor groove-binding architectural proteins: structure, function, and DNA recognition // Annu Rev. Biophys. Biomol. Struct. 1998. Vol. 27. P. 105–131.

25. Huang J., Zhao Y., Liu H., Huang D., Cheng X., Zhao W., Taylor I.A., Liu J., Peng Y.L. Substitution of tryptophan 89 with tyrosine switches the DNA binding mode of PC4 // Scientific Reports. 2015. Vol. 5. P. 8789.

26. Schneider T.D., Stormo G.D., Gold L. Information content of binding sites on nucleotide sequences // J. Mol. Biol. 1986. No 188. P. 415–431.

27. Berg O.G., von Hippel P.H. Selection of DNA binding sites by regulatory proteins—statistical-mechanical theory and application to operators and promoters // J. Mol. Biol. 1987. No 193. P. 723–750.

28. Takeda Y., Sarai A., Rivera V.M. Analysis of the sequence-specific interactions between Cro repressor and operator DNA by systematic base substitution experiments // Proc. Natl. Acad. Sci. USA. 1989. No 86. P. 439–443.

29. Sarai A., Takeda Y. λ repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically // Proc. Natl. Acad. Sci. USA. 1989. No 86. P. 6513–6517.

30. Tanikawa J., Yasukawa T., Enari M., Ogata K., Nishimura Y., Ishii S., Sarai A. Recognition of specific DNA sequences by the c-Myb proto-oncogen product: Role of three repeat units in the DNA-binding domain // Proc. Natl. Acad. Sci. USA. 1993. No 90. P. 9320–9324.

31. Kono H., Sarai A. Structure-based prediction of DNA target sites by regulatory proteins // Proteins. 1999. No 35. P. 114–131.

REFERENCES

1. Si, J., Zhao, R., & Wu, R. (2015). An Overview of the Prediction of Protein DNA-Binding Sites. *International Journal of Molecular Sciences,* 16, 5194-5215. doi: 10.3390/ijms16035194

2. Mandel-Gutfreund, Y., & Margalit, H. (1998). Quantitative parameters for amino acid–base interaction: implications for prediction of protein–DNA binding sites. *Nucleic Acids Research, 26*(10), 2306–2312. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC147552/pdf/262306.pdf

3. Jones, S., Shanahan, H.P., Berman, H.M., & Thornton, J.M. (2003). Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins. *Nucleic Acids Research,* 31, 7189–7198. doi: 10.1093/nar/gfg922

4. Shanahan, H.P., Garcia, M.A., Jones, S., & Thornton, J.M. (2004). Identifying DNA-binding proteins using structural motifs and the electrostatic potential. *Nucleic Acids Research,* 32, 4732–4741. doi: 10.1093/nar/gkh803

5. Ofran, Y., Mysore, V., & Rost, B. (2007). Prediction of DNA-binding residues from sequence. *Bioinformatics, 23*(13), i347–i353. doi: 10.1093/bioinformatics/btm174

6. Luscombe, N.M., Laskowski, R.A., & Thornton, J.M. (2001). Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Research, 29*(13), 2860–2874. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC55782/

7. Lejeune, D., Delsaux, N., Charloteaux, B., Thomas, A., & Brasseur, R. (2005). Protein–Nucleic Acid Recognition: Statistical Analysis of Atomic Interactions and Influence of DNA Structure. *Proteins*, *61*(2), 258–271. doi: 10.1002/prot.20607

8. Tolstorukov, M.Y., Jernigan, R.L., & Zhurkin, V.B. (2004). Protein-DNA hydrophobic recognition in the minor groove is facilitated by sugar switching. *Journal of Molecular Biology,* 337, 65–76. doi: 10.1016/j.jmb.2004.01.011

9. Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., & Mann, R.S. (2010). Origins of specificity in protein-DNA recognition. *Annual review of biochemistry*, 79, 233–26. doi: 10.1146/annurev-biochem-060408-091030

10. Schneider, B., Cerny, J., Svozil, D., Cech, P., Gelly, J.-C., & de Brevern, A. G. (2014). Bioinformatic analysis of the protein/DNA interface. *Nucleic Acids Research, 42*(5), 3381–3394. doi: 10.1093/nar/gkt1273

11. Locasale J.W., Napoli A.A., Chen S., Berman H.M., & Lawson,C.L. (2009). Signatures of protein-DNA recognition in free DNA binding sites. *Journal of Molecular Biology,* 386, 1054–1065. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2753591/

12. Zhao, J., Bacolla, A., Wang, G., & Vasquez, K.M. (2010). Non-B DNA structure-induced genetic instability and evolution. *Cellular and Molecular Life Sciences*, 67, 43–62. doi: 10.1007/s00018-009-0131-2

13. Schneider, B., Neidle, S., & Berman, H. M. (1997). Conformations of the sugar–phosphate backbone in helical DNA crystal structures. *Biopolymers, 42*(1), 113–124. doi:10.1002/(SICI)1097-0282(199707)42:1<113::AID-BIP10>3.0.CO;2-O

14. Varnai, P., Djuranovic, D., Lavery, R., & Hartmann, B. (2002). α-γ transitions in the B-DNA backbone. *Nucleic Acids Research, 30*(24), 5398–5406. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC140057/

15. Zhitnikova, M.Yu., Boryskina, O.P., & Shestopalova, A.V. (2014). Sequence-specific transitions of the torsion angle gamma change the polar-hydrophobic profile of the DNA grooves: implication for indirect protein-DNA recognition. *Journal of Biomolecular Structure and Dynamics, 32*(10), 1670-1685. doi: 10.1080/07391102.2013.830579

16. Berman, H. M., Olson, W. K., Beveridge, D. L., Gelbin, A., Demeny, T., Hsieh, S. H., … Schneider, B. (1992). The nucleic acid database: A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophysical Journal*, 63, 751–759. doi: 10.1016/S0006-3495(92)81649-1

17. Lu, X. J., & Olson, W. K. 3DNA: (2003). A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Research*, 31, 5108–5121. doi: 10.1093/nar/gkg680

18. Ahmad, S., Gromiha, M.M., & Sarai, A. (2004). Analysis and prediction of DNA-binding proteins and their binding residues based on composition, sequence and structural information. *Bioinformatics*, 20, 477–486. doi: 10.1093/bioinformatics/btg432

19. Kuznetsov, I.B., Gou, Z., Li, R., & Hwang, S. (2006). Using evolutionary and structural information to predict DNA-binding sites on DNA-binding proteins. *Proteins,* 64, 19–27. doi: 10.1002/prot.20977

20. Bartlett, G.J., Porter, C.T., Borkakoti, N., & Thornton, J.M. (2002). Analysis of catalytic residues in enzyme active sites. *Journal of Molecular Biology,* 324, 105–121. doi: 10.1016/S0022-2836(02)01036-7

21. Faucher, J., & Pliska, V. (1983). Hydrophobic parameters pi of amino acid side chains from the partitioning of N-acetyl-amino-acid amides. *European Journal of Medicinal Chemistry,* 18, 369–375. Retrieved from https://www.researchgate.net/publication/246404378_Hydrophobic_parameters_II_of_amino_acid_side-chains_from_the_partitioning_of_N-acetyl-amino_acid_amides

22. Kyte, J., & Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology,* 157, 105–132.
Retrieved from http://www.sciencedirect.com/science/article/pii/0022283682905150?via%3Dihub

23. Corona, R. I., & Guo, J.T. (2016). Statistical analysis of structural determinants for protein-DNA binding specificity. *Proteins, 84*(8), 1147–1161. doi: 10.1002/prot.25061

24. Bewley, C.A., Gronenborn, A.M., & Clore, G.M. (1998). Minor groove-binding architectural proteins: structure, function, and DNA recognition. *Annual Review of Biophysics and Biomolecular Structure, 27,* 105–131. doi: 10.1146/annurev.biophys.27.1.105

25. Huang, J., Zhao, Y., Liu, H., Huang, D., Cheng, X., Zhao, W., … Peng, Y.L. (2015). Substitution of tryptophan 89 with tyrosine switches the DNA binding mode of PC4. *Scientific Reports, 5,* 8789. doi: 10.1038/srep08789

26. Schneider, T.D., Stormo, G.D., & Gold, L. (1986). Information content of binding sites on nucleotide sequences. *Journal of Molecular Biology*, 188, 415–431.
Retrieved from http://www.sciencedirect.com/science/article/pii/0022283686901658?via%3Dihub

27. Berg, O.G., & von Hippel, P.H. (1987). Selection of DNA binding sites by regulatory proteins—statistical-mechanical theory and application to operators and promoters. *Journal of Molecular Biology,* 193, 723–750. Retrieved from http://europepmc.org/abstract/med/3612791

28. Takeda, Y., Sarai, A., & Rivera, V.M. (1989). Analysis of the sequence-specific interactions between Cro repressor and operator DNA by systematic base substitution experiments. *Proceedings of the National Academy of Sciences USA*, 86, 439–443. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC286485/

29. Sarai, A., & Takeda, Y. (1989). λ repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically. *Proceedings of the National Academy of Sciences USA*, 86, 6513–6517. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC297874/

30. Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., & Sarai, A. (1993). Recognition of specific DNA sequences by the c-Myb proto-oncogen product: Role of three repeat units in the DNA-binding domain. *Proceedings of the National Academy of Sciences USA*, 90, 9320–9324. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC47559/

31. Kono, H., & Sarai, A. (1999). Structure-based prediction of DNA target sites by regulatory proteins. *Proteins,* 35, 114–131. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6961027