Recognition of transcription factors in complex with operator DNA at atomic resolution

Authors:

Y.N. Chirgadze[1]*, R.V. Polozov[2]

[1] Institute of Protein Research, Russian Academy of Sciences, Pushchino 142290, Moscow Region, Russia

[2] Institute of Theoretical Experimental Biophysics, Russian Academy of Sciences, Pushchino 142290, Moscow Region, Russia

***Corresponding author:**

Y.N. Chirgadze

E-mail: chir@vega.protres.ru

Content of review:

Volume: about 4200 words, 13 Figures, 2 Tables, 15 ref.

Abstract

In this mini-review we have presented our common opinion on one of the key problems of molecular biology and very possibly "molecular medicine". Since the early nineties of last century many significant works have been published on the atomic structure of complexes of transcription factor with DNA-fragments in the form of double-helix, which exist in the cell media. To this time, a few reviews have appeared and we have also finished our study of protein-DNA binding interactions of such complexes. By analyzing numerous significant structural data we have deduced the recognition rules which allow understanding the general feature of such interactions for most wide-spread families of two widespread classes of transcription factors: homeodomains and Zn-fingers. In these cases, we have concentrated only on the main more significant sort of binding DNA-region - its major groove. As the results show, we have deduced several common regularities of these protein-DNA bindings. We have shown that the common type of binding of factor-DNA is determined by the groups of invariant binding contacts while the particular feature of any definite transcription factor is determined by the group of variable contacts.

Keywords: protein-DNA binding, transcription factor, homeodomain, Zn-finger.

1. Introduction

A transcription factor is a protein that binds to a specific sequence of operator DNA, thereby controlling the gene transcription. More than 5% of human genes are known to encode the transcription factors that reflect their biological importance in the cellular functioning. Protein-DNA recognition is an initial key stage, which is followed by several steps. At large distances, the interaction is determined by the electrostatic fields of protein and DNA, which provide for the necessary positions and orientation of the protein along the major groove of the operator DNA. At short interatomic distances, a number of specific contacts are formed between atoms of protein and DNA. Such alternation of expanded analog and accurate digital aspects of binding appears to be a distinctive feature of the recognition process (Chirgadze, et al., 2009). Recognition occurs essentially in the major groove of the DNA molecule through the contacts between amino acid side chains of the protein and bases and phosphates of the DNA molecule. The interaction in the minor groove of DNA usually includes fewer contacts.

To introduce quickly the readers into the molecular situation of spatial structures of the complexes of transcription factors with DNA we have presented examples of such complexes for two widespread families: homeodomain and Zn-C2H2 families (Fig. 1). One can see the common features: modular binding principle, especially for Zn-factor family; and specific binding of single α-helical fragment, which called a recognizing helix as usually. In most cases the recognizing helix is located in the major groove of a double-stranded DNA molecule. Nevertheless, details of recognition for these two families differ significantly, and we can see this below.

2. Polar surface clusters of the transcription factor and their role in binding factors with double-helical RNA and DNA molecule.

The molecular surface of globular proteins are saturated by the polar and charged groups. Eleven of the total twenty amino acid residues contain polar groups and five of them, Asp, Glu, Lys, Arg and His, are charged at neutral pH. On average, there are nearly 30% charged residues in the polypeptide chains of globular proteins.

Detailed analysis of protein surface allows us to discover that polar residues are joined into clusters. Most interesting for functioning and binding are the so-called sing-alternating charge clusters. Originally, they were discovered in the molecule of gamma-crystallin II from calf eye lens and analyzed in the paper of Chirgadze & Tabolina, 1996. This molecule consists of two similar domains, and the surface of each domain displays 2 or 3 sign-alternating clusters of the mentioned types. Six clusters have been found here. An example of such cluster structure is shown in Fig. 2. The structure was obtained in the Laboratory of Prof T. Blundell, University of London, at the rather high resolution of 1.45A. That was enough to observe the structural water molecules and determine accurately their contacts with charged residues. We have also found that charged atoms in such clusters are located very close to the plane of the cluster as seen in Fig 2B. It was shown that the amino acid sequence positions of charged residues are conservative for all the proteins of the gamma-crystallin family of vertebrates including fish, frogs, mice, rats, cattle and humans.

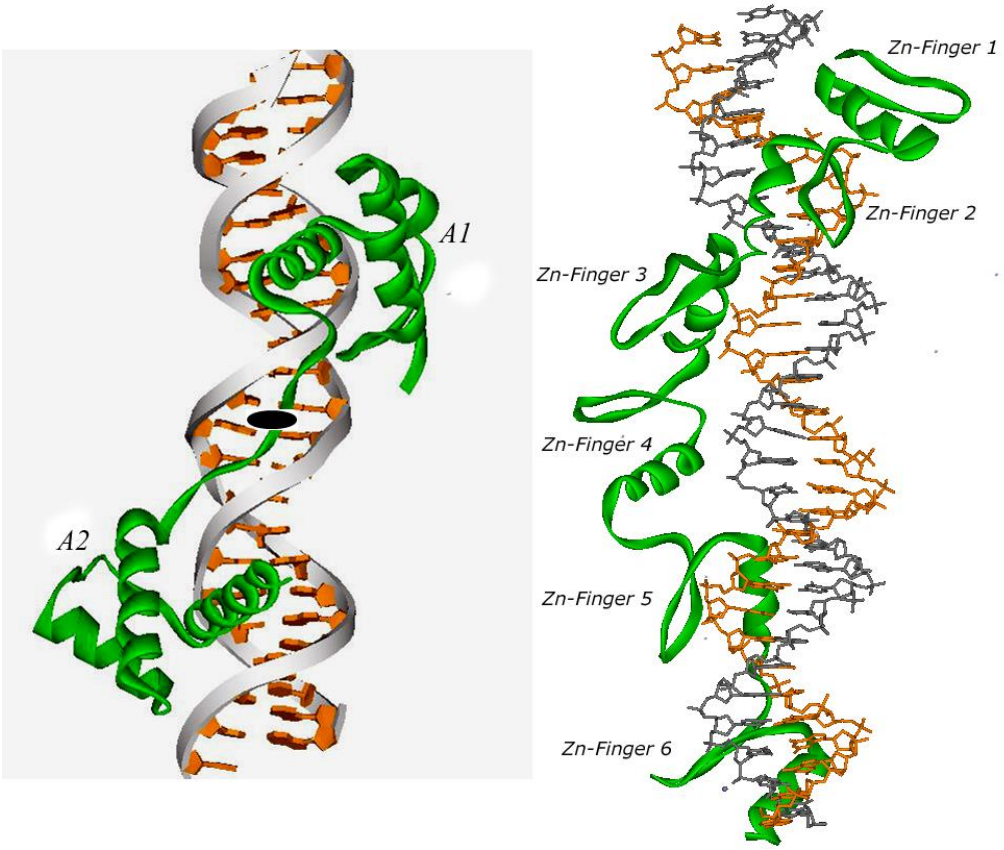

Fig. 1. Crystal structures of two transcription factors belonging to different widespread families. *Left*: Human Pax6 paired domain-DNA complex from homeodomain family (Xu, et al., 1999). The black oval mark designates the position of two-fold pseudo-symmetry axis which shows the centre of the palindrome part of the nucleotide sequence of DNA. *Right*: Frog six-Zn-finger transcription factor IIIA bound with a fragment of the 5S rRNA gene promoter from Zn-C2H2 family (Nolte, et al., 1998).

Large sign-alternating charge clusters formed by the charged side groups of amino acid residues and N- and C-terminal groups were found in the majority of considered globular proteins, namely 235 in a total of 274 selected protein structures, which means 85.8%. The majority of transcription factors also display the presence of charge clusters. The clusters were determined by the criteria proposed earlier: charged groups were included in the cluster if their charged N and O atoms were located at distances between 2.4 and 7.0 Å (Chirgadze & Larionova, 1999). As a result, the alternating charge clusters of proteins should be considered as newly recognized *surface structural invariants*. The importance of the charged side chain clusters is claimed for the updated concept of the protein surface.

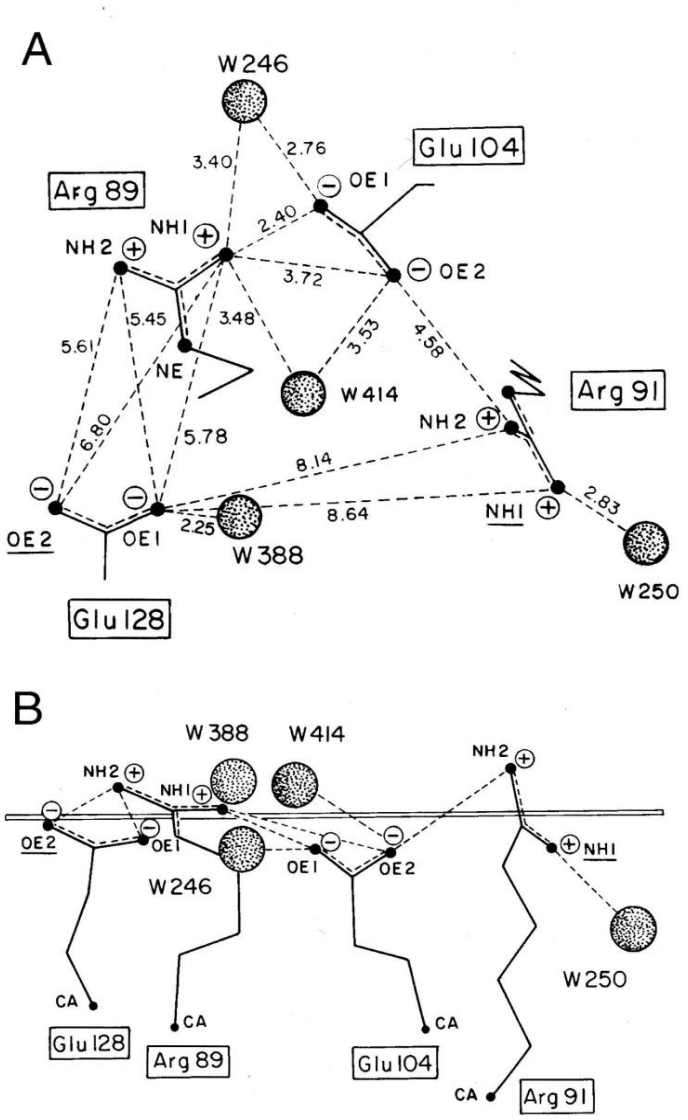

Fig. 2. Two orthogonal views of the skeletal model of charged cluster 89-128 of calf eye lens gamma-crystallin B (Chirgadze & Tabolina, 1996). The distribution of the charged atoms of side chains and water oxygen atoms is shown. The trace of the charge alternating cluster plane is shown by double lines. (A) Plane projection of a cluster. (B) End-view projection of a cluster.

Further investigation has shown the *functional importance* of the large polar clusters which include also charged residues. For example, we have shown that polar clusters determine binding regions of proteins with RNA or DNA. Below we consider in short the results on such complexes. A set of ten non-homologous complexes including twenty independent binding sites was considered of (Chirgadze & Larionova, 2005). Almost all binding polar residues proved to be localized within one largest cluster on the surface of RNA-binding proteins. For instance, 80.5% of binding polar residues have been found within such a cluster in the major RNA grooves of fifteen RNA-binding sites.

A common view of protein-RNA complex for ribosomal protein L25 with double-stranded fragment of 5S rRNA from *E.coli* is presented in Fig. 3. Here we see two different sites with binding polar residues shown as color balls. Cluster analysis of L25 is presented in Fig. 4a. A second example is the complex of viral suppressor of RNA silencing protein p19 from tomato bushy stunt virus and siRNA (Ye & Malinina, 2003). Here, we observed only one binding site on the expanded external beta-sheet structure of p19. Polar binding residues are located in one cluster (Fig. 4b).

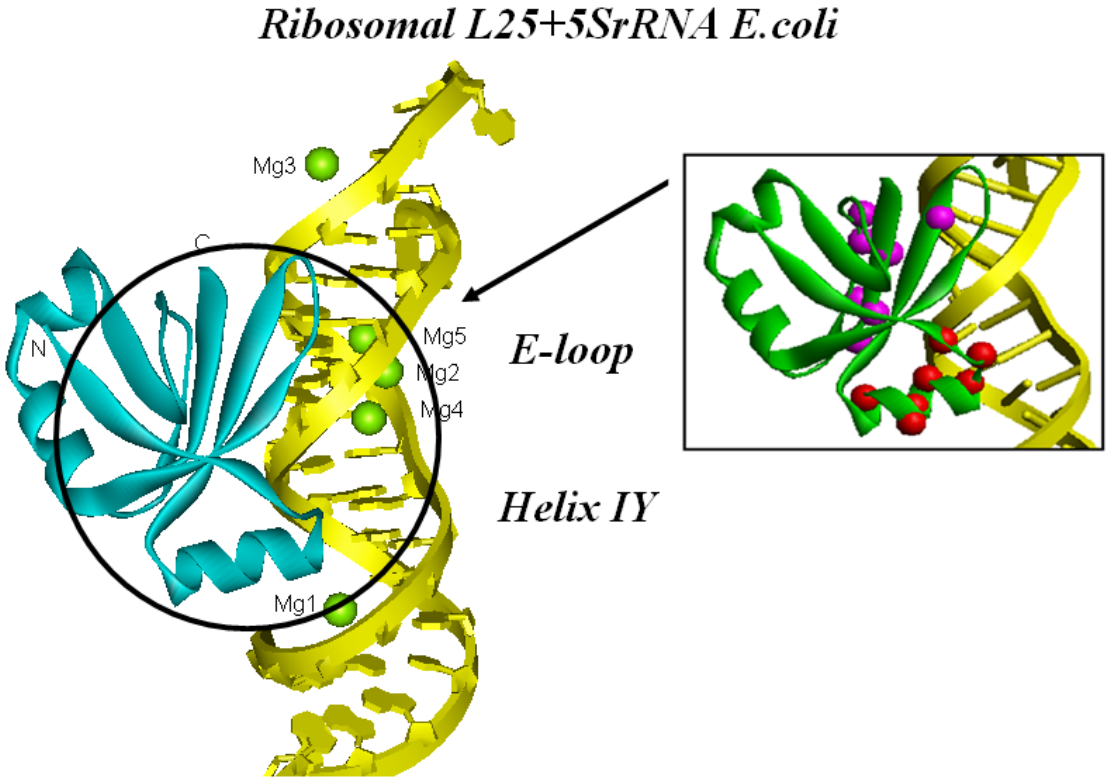

Fig. 3. View of protein-RNA complex for ribosomal protein L25 with double-stranded fragment of 5S rRNA from *E.coli* (Chirgadze & Larionova, 2005). Protein model in right box shows polar residues, as color balls, related to two binding sites.

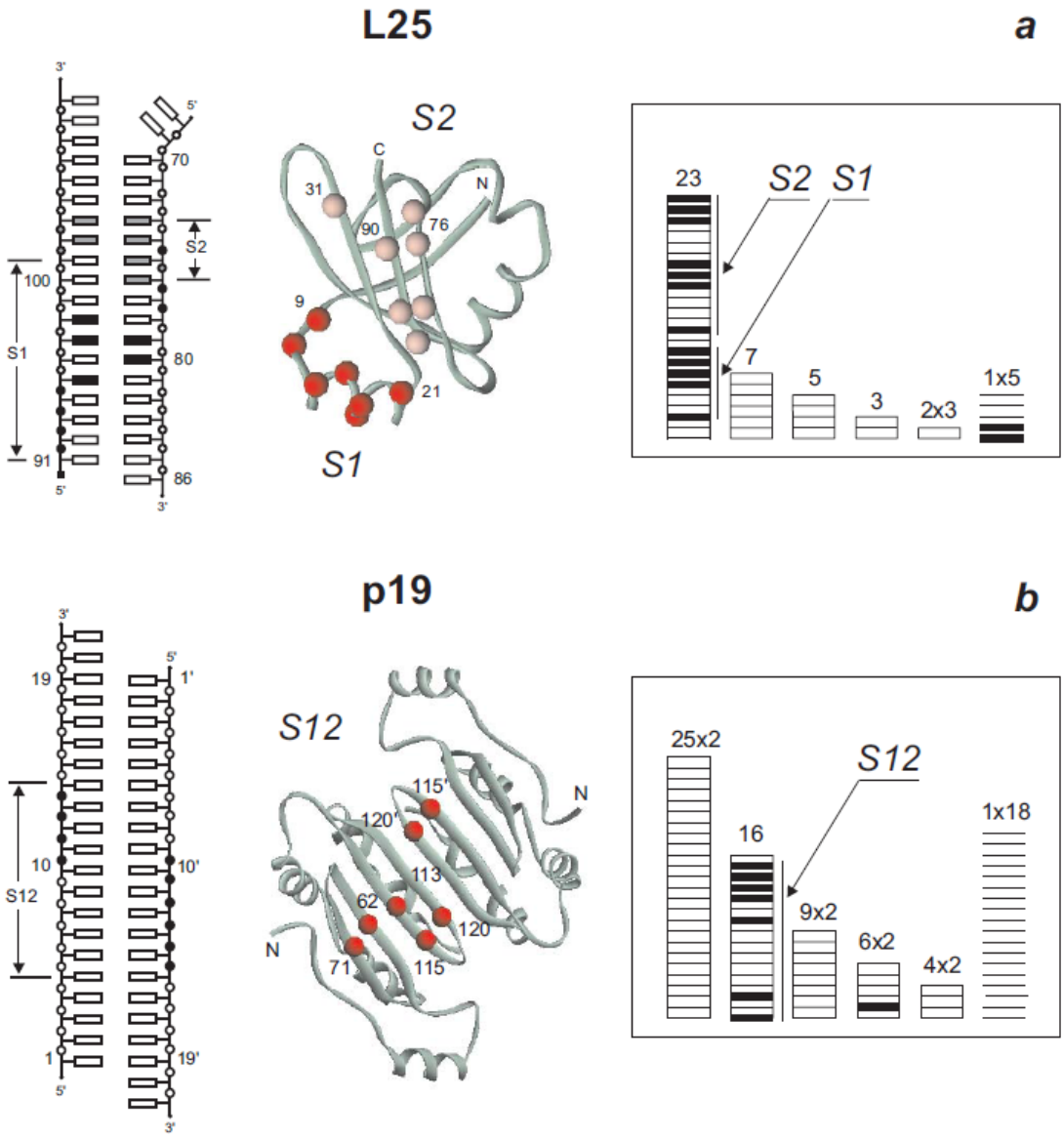

Fig. 4. Atomic contacts in RNA-binding sites: ribosomal protein L25 with 5S rRNA from *E.coli* (Lu & Steitz, 2000) and viral suppressor of RNA silencing p19 from tomato bushy stunt virus and siRNA (Ye & Malinina, 2003).
Left: Schematic RNA fragment, binding sites are marked in black and grey.
Middle: Binding site in the protein models. Two binding sites are shown by a color ball.
Right: Distribution of polar residues in the clusters. Each polar residue corresponds to a thin horizontal line. Contacting polar residues are indicated by solid lines. Each column includes residues of one polar cluster; the first and second numbers above the columns show the number of residues in the cluster and the number of clusters of this size, respectively. The residues are listed in columns according to increasing their number.

3. Polar clusters in binding of transcription factors to operator DNA.

Binding sites of 75 protein domains from 65 complexes of transcription factors with fragments of double-chained operator B-DNA in the major groove region were considered for known spatial structures at the atomic resolution (Chirgadze, Larionova, et al., 2009). The DNA-binding protein domains belong to various structural families and differ greatly in the molecular structure and size. However, in all cases mainly polar residues of *the only recognizing α-helix* of protein factor ensured the binding as shown in Fig. 5. We have found that binding helical residues have a very clear novel recognition sign. That means: for 95.8% of total protein domains the binding polar residues of the recognizing α-helix were localized within one of the two largest clusters of polar side groups on the protein surface. Examples of ten different transcription factor-DNA complexes with clusters of binding polar groups are presented in Fig. 6.

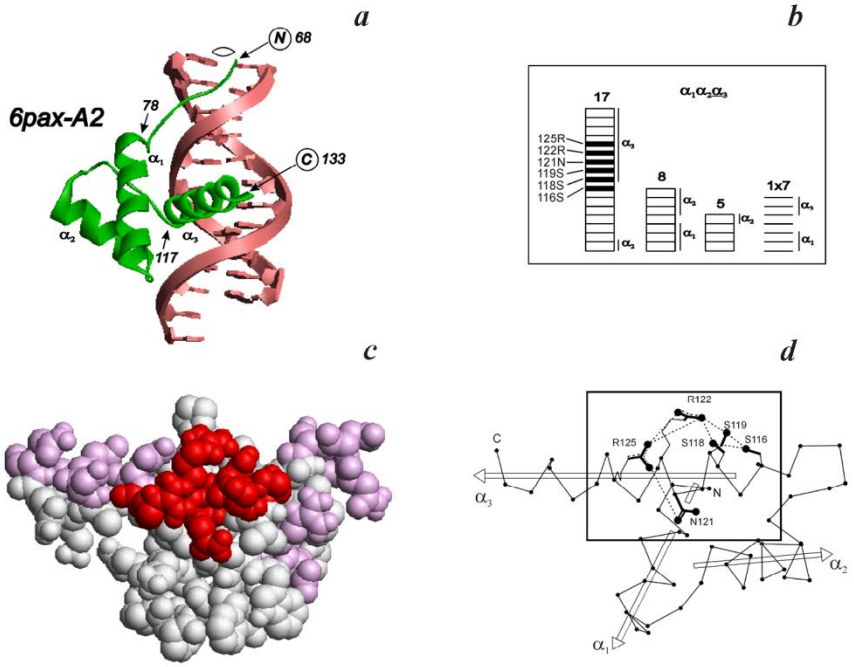

Fig. 5. Binding complex 6PAX of transcription factor with a fragment of B-DNA (Xu *et al*., 1999): *a* - Spatial model of domain A2 of transcription factor with a fragment of B-DNA. *b* - Distribution of polar side groups of residues among clusters on the protein surface, *c* - Arrangement of DNA-binding polar residues (red color sub-cluster) and the rest polar residues of the largest cluster (pink color) on the surface of a protein domain. *d* - Formation of a cluster by polar groups at interatomic distances between 4.0 and 7.0 Å.

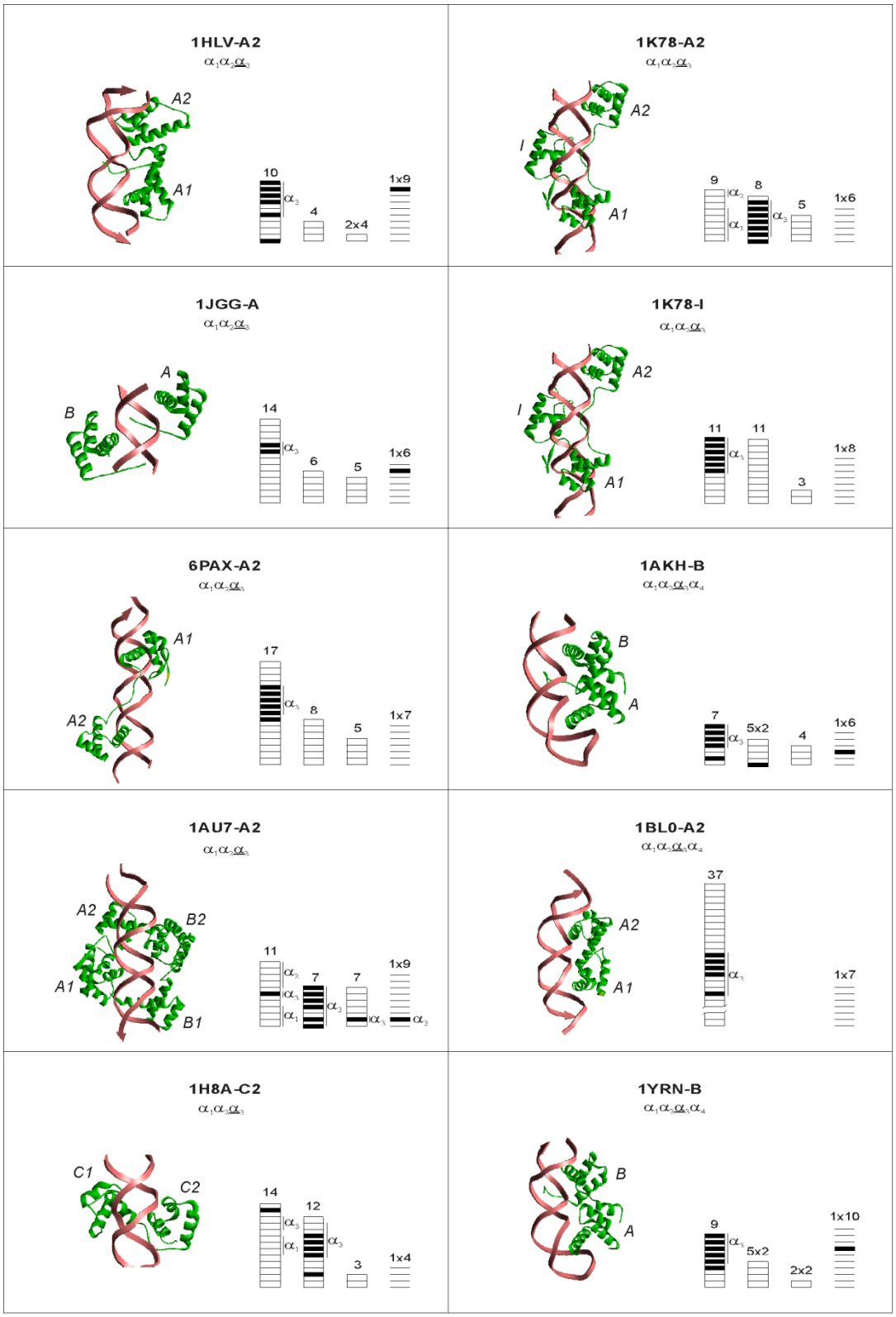

Fig. 6. Examples of ten transcription factor-DNA complexes with clusters of binding polar groups. Within each box: left side: Atomic model of the complexes. Right side: Distribution of residues in polar clusters, where the binding residues are shown with thick horizontal rods.

4. Recognition rules for binding of homeodomain transcription factors to operator DNA.

Among transcription factors, homeodomains are nearly the most important in eukaryotes since they control differentiation development of embryonic cells into tissue or organ-specific cells. Initially discovered in fruit flies, these factors have been found in all vertebrates (Cappen, 2005). Structurally, homeodomains are three-helical bundles accompanied at their N-terminus by a basic loop. Their recognizing helix is the third C-terminal helix, and this helix is penetrating the major groove of DNA. Interactions of homeodomains with DNA have been extensively reviewed in (Ledneva et al., 2001).

Let us consider the functional meaning of binding contacts and classify them. We have carried out this for the spatial structure of transcription factor with a fragment of operator DNA of homeodomain Msx-1 from a mouse *Mus musculus* (Hovde et al., 2001). The simplified structure of the complex is presented in Fig. 7. In accordance with the general conclusion of Point 3 of this review one can see the DNA binding domain consisting of three α-helices which form binding contacts of only third recognizing helix with the region of major groove of double-stranded B-DNA helix. On the basis of this particular complex we have considered general binding regularities of the whole homeodomain complexes with operator DNA (Chirgadze, Zheltukhin et al., 2009). All contacts can be divided into several groups which are related with different functional sense. A description of these groups is presented in the legend to Fig. 8, where they are shown in various colors. Thus, a rather complicated cluster of homeodomain contacts becomes more understandable and clear.

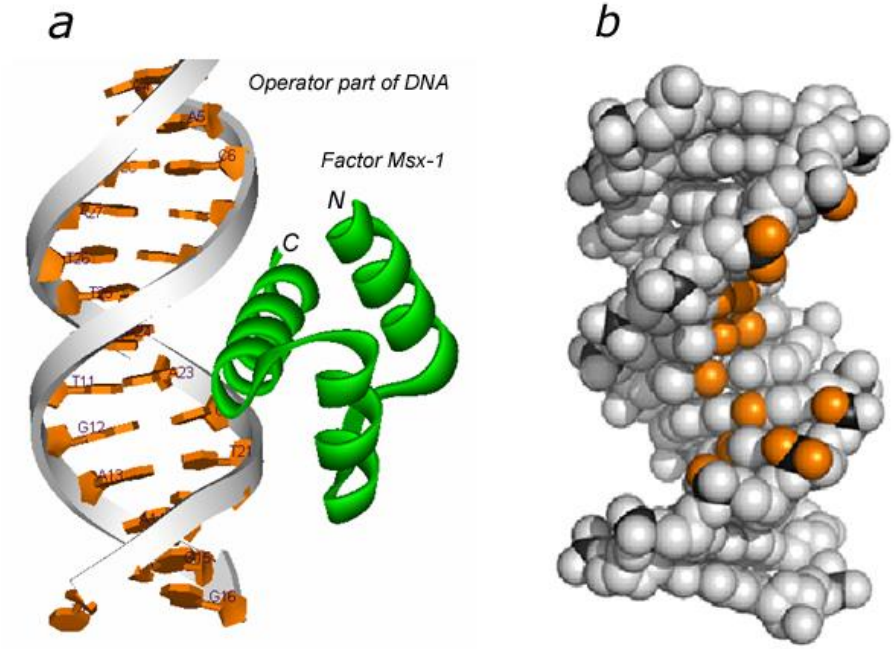

Fig. 7. Spatial models of transcription factor and fragment of operator DNA of homeodomain Msx-1 from a mouse *Mus musculus* (Hovde et al., 2001): *a* - Ribbon model of DNA- factor complex. Only one of total two similar protein domains is shown. The third recognizing α-helix forms contacts with DNA in the region of its major groove; *b* - Space-filling model of DNA. Nucleotide atomic groups, which can contact with transcription factor, are marked in light brown.

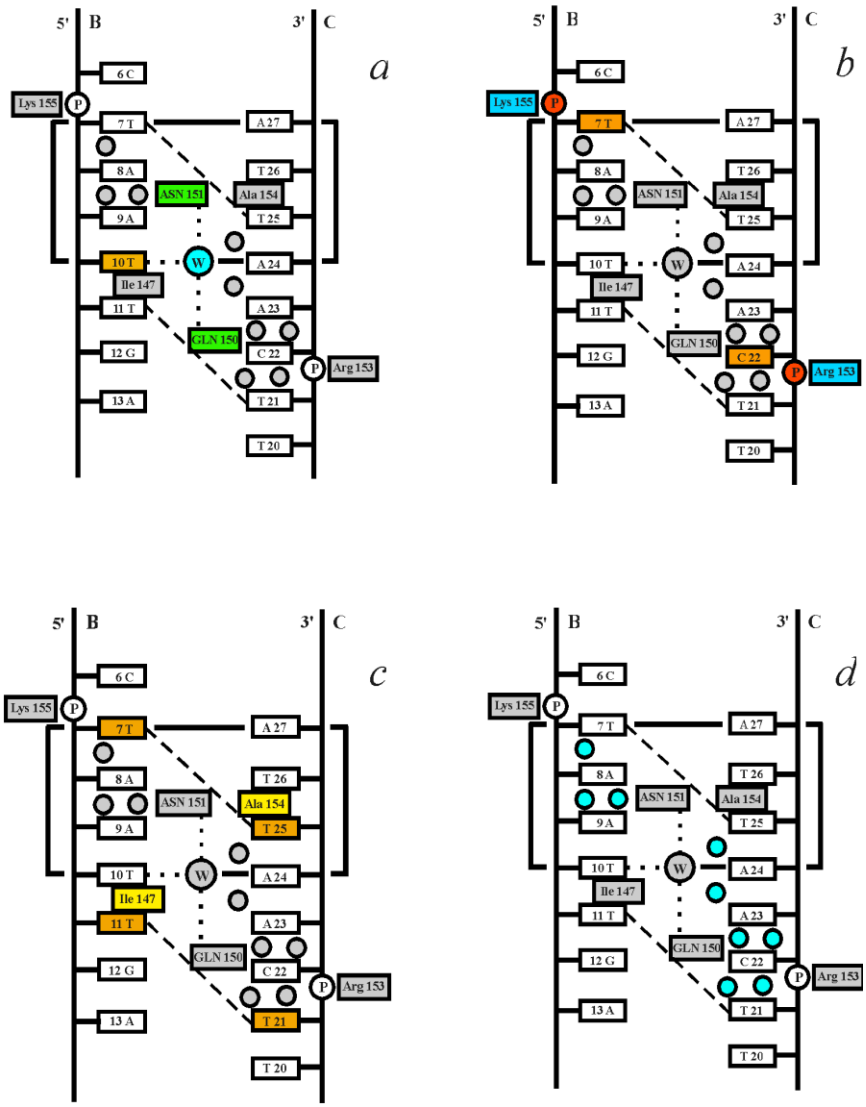

Fig. 8. Specific functional senses of different homeodomain-DNA contacts as considered for transcription factor homeodomain Msx-1. Marked in color: (***a***) Protein and DNA atomic groups bound unique structural water molecule, (***b***) Positively charged compensators of phosphate negative charges of both DNA chains, (***c***) Nonpolar fixing contact barriers of the protein-DNA binding site, (***d***) Set of 10-20 water molecule nets providing local conformational mobility.

Consider now all contacts occurring in the representative set of twenty two atomic structures of yet known homeodomain-DNA complexes (Chirgadze, et al., 2012). Two groups of invariant and variable contacts have been observed. The main focus of analysis was devoted to the invariant contacts. We have found a certain position-specific set of invariant contacts with high occurrence frequencies which was present in all structures of homeodomain-DNA complexes. Remarkably, this set of contacts was evolutionary conserved for different taxonomic groups of the homeodomain

family. It is comprised of some highly conservative DNA-protein contacts: adenine-asparagine contact and several position-specific phosphate contacts with the positively charged amino acid residues lysine and arginine. We have assumed that these invariant contacts will be considered as a specific recognition rule for complex formations of homeodomains with the operator DNA. The result can be presented in the form of a molecular model and binding contact diagram as shown in Fig. 9.

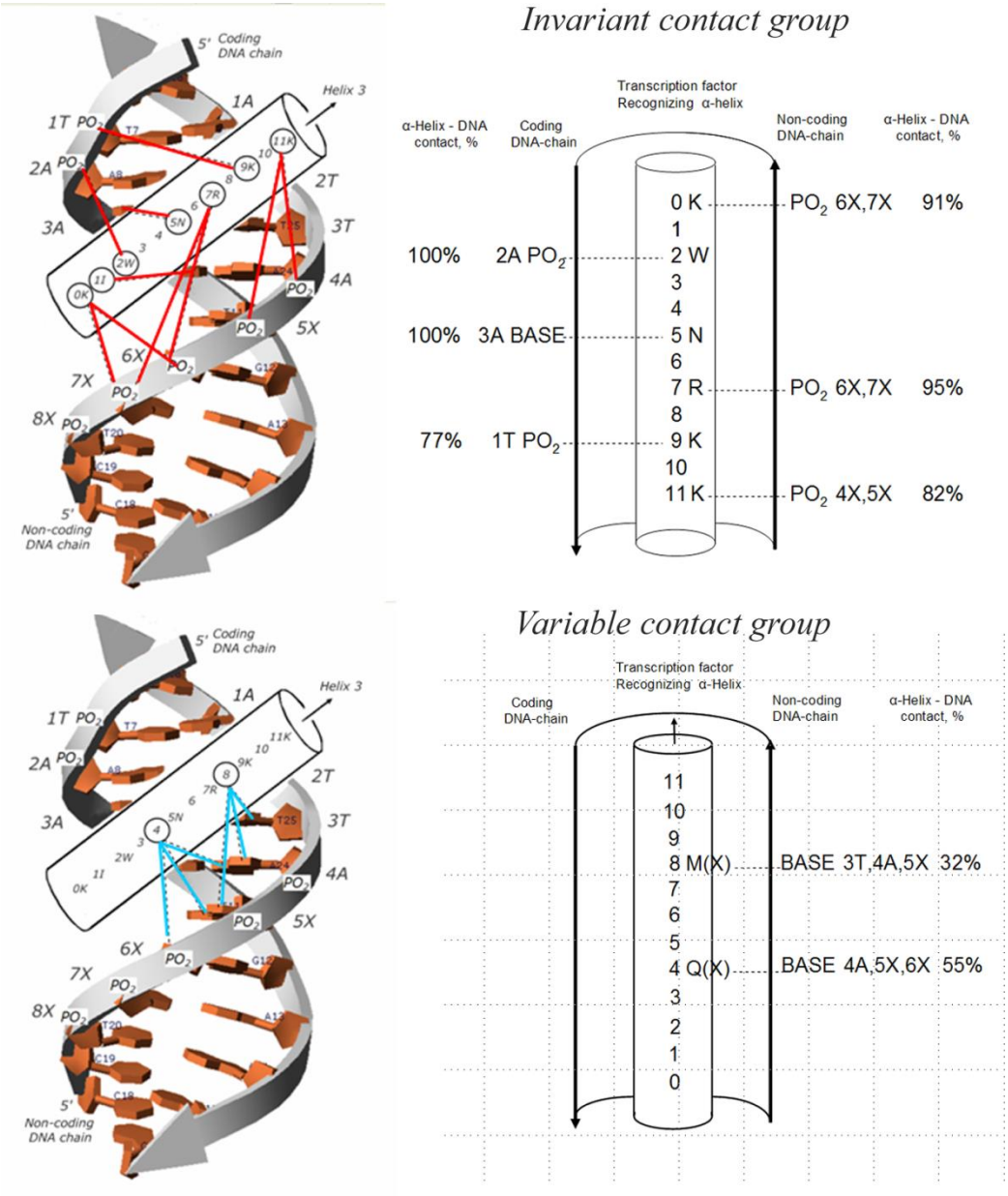

Fig. 9. Invariant (upper case) and variable (lower case) groups of homeodomain-DNA contacts are shown on molecular models and simple structural diagrams of contacts. All considered contacts are formed by recognizing helix situated at the major groove of double-stranded DNA.

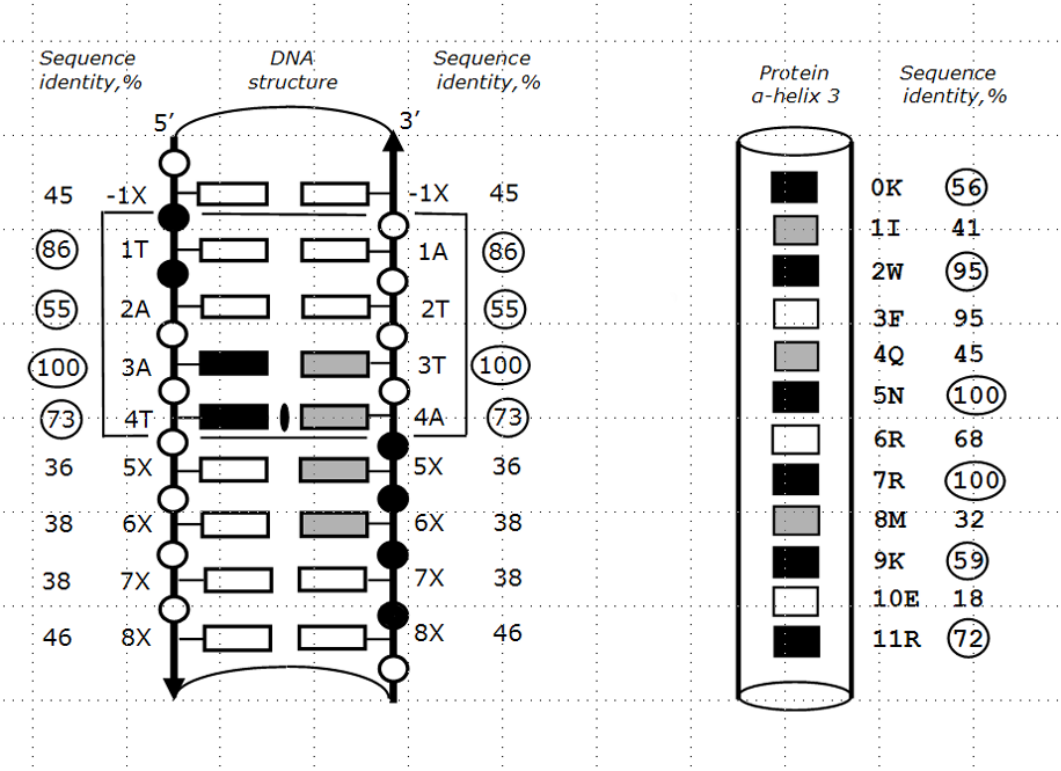

Fig. 10. Recognition patterns of the double-stranded operator DNA (left) and the recognizing α-helix of homeodomain factor (right). All contacting phosphate and nucleotide bases groups of DNA and side groups of protein recognizing α-helix are shaded: groups of invariant contacts are in black, groups of variable contacts are in grey, non-binding groups are left in white. The canonical part of the DNA promoter region is designated by a rectangle box. Significant numbers of sequence identity higher than 55% are shown by circles. The internal numbering systems were introduced within the recognizing α-helix and for two DNA chains.

A structural diagram of considered recognition patterns is presented in general form in Fig. 10. We can look here at contact group recognition patterns separately, both of DNA and protein recognizing helix. Here, all contacts are divided into two groups: invariant contacts (black colored) and variable contacts (grey colored). Non-contacting atomic groups are shown as white boxes. It is important to emphasize that these recognition patterns are common for the whole set of considered homeodomain complexes with DNA. Moreover, we have further confirmed that this contact pattern is inherent solely to the homeodomain family. We can pay special attention to invariant contacts of the homologous canonical part $T_1A_2A_3T_4$ of the coding DNA chain, and contacts of the variable sequence of the non-homologous part $A_4X_5X_6X_7$ of the non-coding DNA chain. In the recognizing helix, the majority of invariant contacts are concentrated at N- and C-terminal ends which form contacts with the phosphate groups of DNA, with one exception – asparagine N5. In contrast, two variable contacts of the recognizing helix both are located in the middle of the helix.

5. Recognition rules for binding of Zn-Cys2His2 transcription factors to operator DNA.

The molecules of Zn-finger transcription factors consist of several similar canonical protein units. We analyzed the crystal structures for 46 such basic units from a total of twenty two complexes of Zn-Cys2His2 family with the fragments of operator DNA (Polozov, et al., 2015). In all cases, their recognizing α-helices are located in the major groove of DNA and form specific invariant contacts with the nucleotide or phosphate groups of DNA.

First of all, we must take into account the specific modular type of binding of transcription factor's units with double-stranded DNA (Fig. 11). The additional main feature is the overlapping of binding contacts of each basic unit with adjacent DNA-triplet. That means one basic unit contact with tetra-nucleotide ZXYZ where the first nucleotide Z belongs to the adjacent next DNA-triplet. And this remarkable contact is formed with invariant residue histidine 7 of recognizing α-helix.

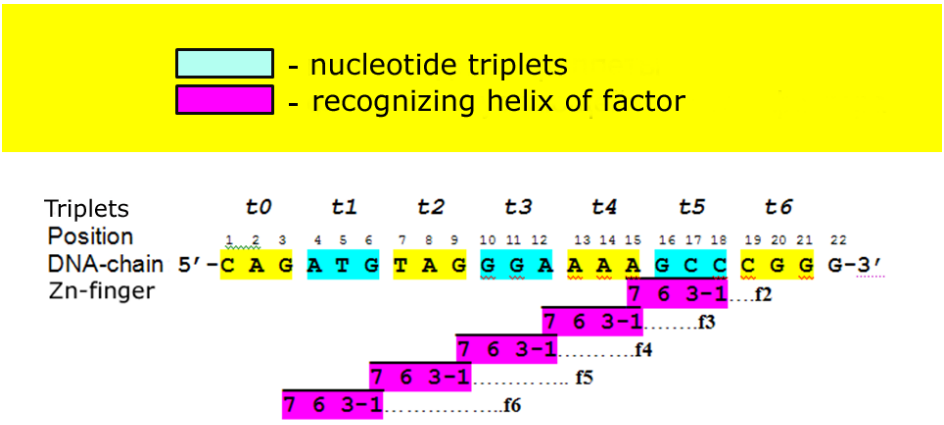

Fig. 11. Modular principle of recognition of complexes of transcription factors belonging to the Zn-Cys2Hys2 family with the coding DNA chain. The DNA recognition site consists of consecutive nucleotide triplets. Each Zn-finger unit contacts DNA molecule by four residues of recognizing α-helix. The residues of helix were enumerated according to the internal numbering.

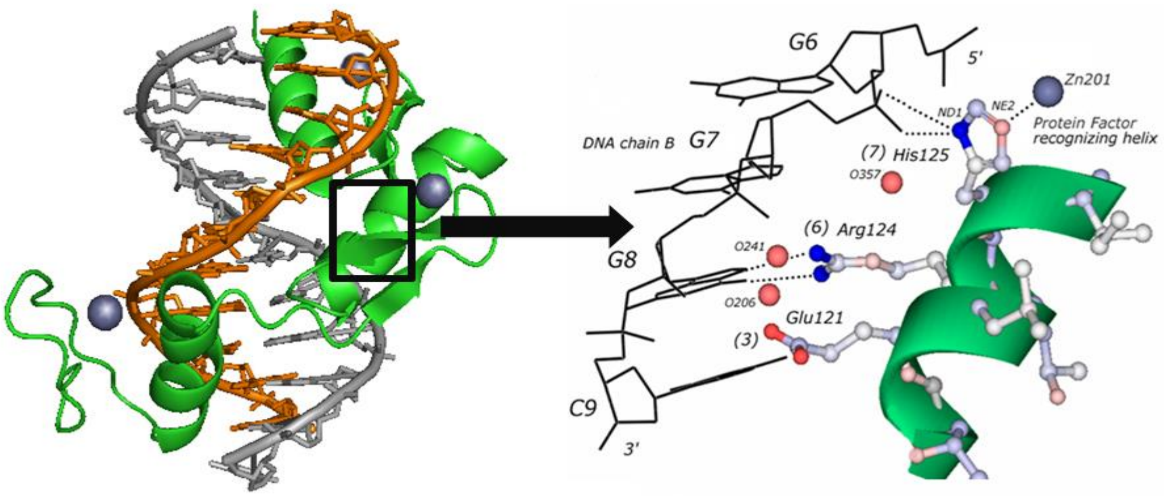

Fig. 12. Atomic structure of the transcription factor Zif268 complex with a fragment of operator DNA from mouse Mus musculus (Elrod-Erickson et al., 1996; PDB code 1aay). *Left*: Complex contains three nucleotide triplet units and three Zn-finger protein units. The coding DNA chain is painted in orange color and ion Zn is shown as a large grey ball. *Right:* High resolution model shows protein-DNA contacts, small red balls - water molecules.

Let us consider, as an example, the structure of the well-characterized mouse Zif268 complexes, which contained three Zn-finger units (Elrod-Erickson, et al., 1996). The structure of this complex is presented in Fig. 12. The protein consists of 90 residues and includes three basic units with three Zn-ions. Each protein factor unit has structure Zn-ββα and called Zn-finger (Table 1).

Table 1. Binding contacts of operator DNA and recognizing factor helix of three basic Zn-fingers in Zif268 transcription factor belonging to the Zn-Cys2His2 family.

Coding DNA chain

```
Zif268    DNA tripet                                            Amino acid position
PDB code underlined                                     -5-4-3-2-1 1 2 3 4 5 6 7 8 9 10 11 12
                              β β β                      β β β β t α α α α α α α α α α α α

Zn-fingers  * * * *   1                 10              *   20  *   * *          30
1aay Af1    G G C G   M E R P Y A C P V E S C D R R F S R S D E L T R H I R I H T G Q K    33
1aay Af2    G T G G         P F Q C R I - - C M R N F S R S D H L T T H I R T H T G E K    61
1aay Af3    G G C G         P F A C D I - - C G R K F A R S D E R K R H T K I H L R Q K D  90
```

Notes. Color coding of contact types: yellow - amino acids with phosphates, and cyan - amino acids with nucleotide bases. Asterisks and red colors in the numbering designate protein-DNA contacts as observed in the whole Zn-Cys2His2 family.

We have shown that recognition of DNA occurs via five protein contacts. The canonical binding positions of the recognizing α-helix residues were -1, 3, 6, and 7, which make contacts with the tetra-nucleotide sequence ZXYZ of the coding DNA strand. Here, the canonical binding triplet is XYZ. The non-coding DNA strand forms only one contact at α-helix position 2. For the coding DNA chain only one highly conservative contact His7α has been observed with the phosphate group of nucleotide Z which proceeds to triplet XYZ (Table 2). This particular contact is invariant for all transcription factors belonging to the Zn-Cys2His2 family. It had high frequency of occurrence of 83%, which we considered as an invariant recognition rule for this family. Details of these contacts for the high resolution structure of mouse factor Zif268 can be seen in Fig. 10. Recognition patterns of binding for operator DNA and recognizing α-helix of Zn-Cys2His2 transcription factor is presented in Fig. 13.

Table 2. Contacts of operator DNA and recognizing protein helix in the Zn-Cys2His2 family transcription factors.

Color coding of contacts: yellow – amino acids with phosphates, and cyan – amino acids with bases.

Other designations see in Table 1.

Coding DNA chain

Complex PDB code	DNA tripet underlined	Amino acid position -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10 11 12
		β β β β t α α α α α α α α α α α α
Zif268	* * * *	* * * *
1aay Af1	G G C G	R R F S R S D E L T R H I R I H T
1aay Af2	G T G G	R N F S R S D H L T T H I R T H T
1aay Af3	G G C G	R K F A R S D E R K R H T K I H L
Zif268 var1		
1a1k Af1	G G A C	R R F S R S A D L T R H I R I H T
1a1f Af1	G G A C	R R F S D S S N L T R H I R I H T
1a1g Af1	G G C G	R R F S D S S N L T R H I R I H T
1a1h Af1	G G C A	R R S F Q S G S L T R H I R I H T
1a1i Af1	G G C A	R R F S R S A D L T R H I R I H T
1a1l Af1	G G C A	R R F S R S D E L T R H I R I H T
1a1j Af1	G G C G	R R F S R S A D L T R H I R I H T
Zif268 TATA box		
1g2d Cf1	A A A A	R R F S Q K T N L D T H I R I H T
1g2d Cf2	T A T A	R N F S Q H T G L N Q H I R T H T
1g2d Cf3	C G C T	R K F A T L H T R D R H T K I H L
1g2f Cf2	T A T A	R N F S Q Q A S L N A H I R T H T
Zif268 var2		
1jk1 Af1	G G C G	R R F S R S A E L T R H I R I H T
1jk2 Af1	G G C T	R R F S R S A E L T R H I R I H T
Zif23-GCN4		
1llm Df2	G T G G	R N F S R S D H L T T H I R T H T
1llm Df3	C G C G	R K F A R S D E R K R H R D T I Q
1UBD		
1ubd Cf1	A G A C	K M F R D N S A M R K H L H T H G
1ubd Cf2	T G G A	K A F V E S S K L K R H Q L V H T
1ubd Cf3	A A A T	K R F S L D F N L R T H V R I H T
1ubd Cf4	T C A A	K K F A Q S T N L K S H I L T H A
Aart		
2i13 Af1	C C G G	K S F S R S D H L A E H Q R T H T
2i13 Af2	A G C C	K S F S D K K D L T R H Q R T H T
2i13 Af3	A A A A	K S F S Q R A N L R A H Q R T H T
2i13 Af4	G G G A	K S F S Q L A H L R A H Q R T H T

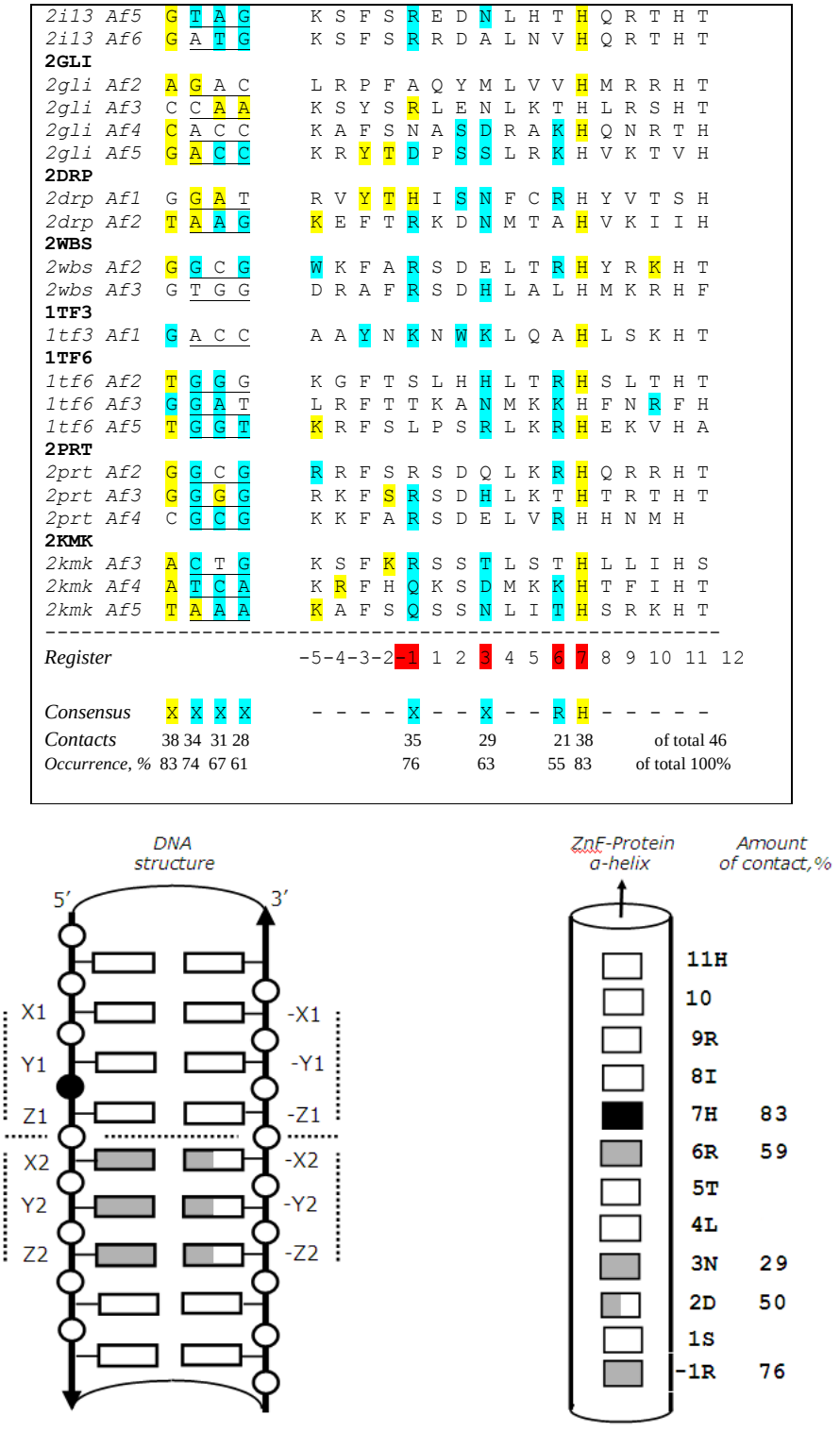

Fig. 13. Recognition patterns of the double-stranded operator DNA with two triplets (left) and one recognizing α-helix of Zn-Cys2His2 transcription factor (right). All contacting phosphate and bases groups of DNA and side groups of recognizing α-helix are colored: invariant contacts are in black, variable contacts are in grey, non-binding groups are in white. Alternatively bound helix position 2 is shown as grey/white box.

6. Conclusion

We can see now that two main principles of transcription factor - DNA recognition are realized:

- binding region of transcription factors is related with their polar surface areas,

- binding in the major grooves of DNA occurs by recognizing α-helix of transcription factors.

One can see that at the atomic level the recognition is rather complicated and we would say even more sophisticated. Of course there are reasons for this. And one important of them is participation of structural key water molecules, as we can see considering that in the complex of homeodomain family. Some important and interesting details, and similarity and differences of two considered family transcription families are presented in the Table 3.

Table 3. Comparison of complexes of transcription factors of homeodomain and Zn-Cys2His2 family with operator DNA .

Feature	Homeodomains	Zn-Cys2His2
DNA molecule	Double-chained helix	Double-chained helix
DNA binding site	Major grove of the molecule	Major grove of the molecule
DNA recognition site	Total 7-8 nucleotide pairs	Total 6-27 nucleotide pairs, one site includes a few triplet units
DNA sequence site	XYZ-TAAT-XYZ	XYZ-XYZ - any nucleotide
Protein molecule	Protein 3α peptide unit	Zn-finger βα peptide unit
Protein binding site	Recognizing α-helix total 12 residues	Recognizing α-helix - 12 residues and β-strand - 1 residue
Protein site sequence	10 residues of total 12	5 residues of total 13
Invariant contacts	Two contacts - coding DNA chain One contact - non-coding DNA chain	One contact - coding DNA chain
Variable contacts	Two contacts - coding DNA chain Two contacts with non-coding DNA chain	Three contacts - coding DNA chain One contact with non-coding DNA chain
Recognition rules	One contact Asn-Water-Ade Six contacts of α-helix with phosphates of DNA	One contact of α-helix His7 with phosphate of DNA

Finally, we must underline the essential feature of binding contacts. There are two groups of contacts: invariant and variable. The first one determines invariant binding of a given family type. And the second is related to binding of any definite transcription factor. The last one has an important significance for medicine when the very efficient, but maybe not safe at the present time, so-called gene therapy is going to be used.

7. Acknowledgement

The author would like to express their gratitude to Dr Nickolay Chirgadze, X-CHIP Technologies Inc., for the valuable discussion of the review concept and the promotion for the publication of this review.

References

Cappen C. In: *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics, Part 1.1. Genetic Variation and Evolution*. DOI: 10.1002/047001153X.g101209 (2005) http://mrw.interscience.wiley.com/ggpb/articles/g101209/frame.html

Chirgadze Yu.N., Tabolina O.Yu. Alternating charge clusters of side-chains: new surface structural invariants observed in calf eye lens gamma-crystallins. *Protein Engineering*, 1996, 9, 745-754.

Chirgadze Yu.N., Larionova E.A. Spatial sign-alternating charge clusters in globular proteins. *Protein Engineering*, 1999, 12, 101-105.

Yu. N. Chirgadze and E. A. Larionova, The key role of charge clusters of polar residues of RNA-binding proteins in the formation of complexes with RNA. *Mol. Biologia (English transl.)*, 2005, 39, 892-905.

Yu. N. Chirgadze, E. I. Zheltukhin, R. V. Polozov, V.S. Sivozhelezov , V.V. Ivanov. Binding regularities in complexes of transcription factors with operator DNA: homeodomain family. *J. Biomol. Structure & Dynamics*, 2009, 26, 687- 700.

Chirgadze, E.A. Larionova, V.V. Ivanov. Novel Recognition Sign of DNA-binding alpha-Helix in Complexes of Transcription Factors from Different Families with Operator DNA, *J. Biomol. Structure & Dynamics*, 2009, 27, 83-96.

Chirgadze Yu. N., Sivozhelezov V. S., Polozov R.V., Stepanenko V.A., Ivanov V.V. Recognition Rules for Binding of Homeodomains to Operator DNA, *J. Biomol. Structure & Dynamics*, 2012, 29, 715-731.

Elrod-Erickson, M., Rould, M. A., Nekludova, L., & Pabo, C. O. (1996). Zif268 protein-DNA complexes refined at 1.6A: A model system for understanding zinc finger-DNA interactions. *Structure*, 4, 1171-1180.

Hovde S., Abate-Shen, C, and Geiger, J.H. (2001). Crystal structure of the Msx-1 homeodomain-DNA complex. *Biochemistry*, *40*, 12013-12021.

Ledneva, R.K., Alekseevskii, A.V., Vasiliev, S.A., Spirin, S.A., Kariagina, A.S. (2001). Structural aspects of homeodomain interactions with DNA. *Mol. Biol. (Moscow)*, 35, 764-777.

Lu Min, Steitz T.A. 2000. Structure of *E. coli* ribosomal protein L25 complexed with a 5S rRNA fragment at 1.8Å resolution. *Proc. Natl. Acad. Sci. USA*. **97**, 2023-2028.

Nolte, R. T., Conlin, R. M., Harrison, S. C., & Brown, R. S. (1998). Differing roles for zinc fingers in DNA recognition: Structure of a six-finger transcription factor IIIA complex. *Proc. Natl. Acad. Sci. USA*, *95*, 2938-2943.

Polozov R.V., Sivozhelezov V.S., Chirgadze Yu.N., Ivanov V.V. Recognition rules for binding of Zn-Cys2His2 transcription factors to operator DNA. 2015, *J. Biomol. Structure & Dynamics*, 33, 253-266.

Xu He, Rould MA, Xu W, Epstein JA, Maas RL, Pabo CO. Crystal structure of the human Pax6 paired domain–DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes & Development*. 1999;13(10):1263-1275.

Ye Keqiong, Malinina L., Patel D.J. 2003. Recognition of small interfering RNA by a viral suppressor of RNA silencing. *Nature*. **426**, 874–878.

Xu H.E., M. A. Rould, W. Xu, J. A. Epstein, R. L. Maas, and C. O. Pabo.1999. *Genes & Development*, *13*, 1263-1275.

