Journal of Biomolecular Structure and Dynamics

ISSN: 0739-1102 (Print) 1538-0254 (Online) Journal homepage: http://www.tandfonline.com/loi/tbsd20

The electrostatic role of the Zn-Cys2His2 complex in binding of operator DNA with transcription factors: mouse EGR-1 from the Cys2His2 family

Y.N. Chirgadze, E.A. Boshkova, R.V. Polozov, V.S. Sivozhelezov, A.V. Dzyabchenko, M.B. Kuzminsky, V.A. Stepanenko & V.V. Ivanov

To cite this article: Y.N. Chirgadze, E.A. Boshkova, R.V. Polozov, V.S. Sivozhelezov, A.V. Dzyabchenko, M.B. Kuzminsky, V.A. Stepanenko & V.V. Ivanov (2018) The electrostatic role of the Zn-Cys2His2 complex in binding of operator DNA with transcription factors: mouse EGR-1 from the Cys2His2 family, Journal of Biomolecular Structure and Dynamics, 36:15, 3902-3915, DOI: 10.1080/07391102.2017.1404937

To link to this article: https://doi.org/10.1080/07391102.2017.1404937

Accepted author version posted online: 13 Nov 2017.
Published online: 07 Jan 2018.

Journal of Biomolecular Structure and Dynamics, 2018
Vol. 36, No. 15, 3902–3915, https://doi.org/10.1080/07391102.2017.1404937

The electrostatic role of the Zn-Cys2His2 complex in binding of operator DNA with transcription factors: mouse EGR-1 from the Cys2His2 family

Y.N. Chirgadze[a]*, E.A. Boshkova[a], R.V. Polozov[b], V.S. Sivozhelezov[c], A.V. Dzyabchenko[d], M.B. Kuzminsky[e], V.A. Stepanenko[f] and V.V. Ivanov[f,g]

[a]*Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow Region 142290, Russia;* [b]*Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, Moscow Region 142290, Russia;* [c]*Institute of Cell Biophysics, Russian Academy of Sciences, Pushchino, Moscow Region 142290, Russia;* [d]*Karpov Institute of Physical Chemistry, State Research Center, Vorontsovo pole 10, Moscow 105064, Russia;* [e]*Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninsky prospect 47, Moscow 119991, Russia;* [f]*Joint Institute for Nuclear Research, Dubna, Moscow Region 141980, Russia;* [g]*National Research Nuclear University MEPhI, Moscow, 115409, Russia*

Communicated by Ramaswamy H. Sarma

(*Received 9 September 2017; accepted 1 November 2017*)

The mouse factor Zif268, known also as early growth response protein EGR-1, is a classical representative for the Cys2His2 transcription factor family. It is required for binding the RNA polymerase with operator dsDNA to initialize the transcription process. We have shown that only in this family of total six Zn-finger protein families the Zn complex plays a significant role in the protein-DNA binding. Electrostatic feature of this complex in the binding of factor Zif268 from *Mus musculus* with operator DNA has been considered. The factor consists of three similar Zn-finger units which bind with triplets of coding DNA. Essential contacts of the factor with the DNA phosphates are formed by three conservative His residues, one in each finger. We describe here the results of calculations of the electrostatic potentials for the Zn-Cys2His2 complex, Zn-finger unit 1, and the whole transcription factor. The potential of Zif268 has a positive area on the factor surface, and it corresponds exactly to the binding sites of each of Zn-finger units. The main part of these areas is determined by conservative His residues, which form contacts with the DNA phosphate groups. Our result shows that the electrostatic positive potential of this histidine residue is enhanced due to the Zn complex. The other contacts of the Zn-finger with DNA are related to nucleotide bases, and they are responsible for the sequence-specific binding with DNA. This result may be extended to all other members of the Cys2His2 transcription factor family.

Keywords: transcription factor; Zn-finger; Zif268; EGR-1

1. Introduction

Transcription factors of the Cys2His2 family are the most abundant DNA-binding motifs found in eukaryotes. It was discovered as factor IIIA in frog oocytes from *Xenopus laevis* (Miller, McLachlan, & Klug, 1985). Further, the atomic structure of factor Zif268 from *Mus musculus* was solved and used for various structural and functional analyses (Elrod-Erickson, Benson, & Pabo, 1998; Elrod-Erickson, Rould, Nekludova, & Pabo, 1996; Pavletich & Pabo, 1991). The factor includes three Zn-finger protein units which bind to the fragment of dsDNA with three nucleotide triplets. Details of interatomic contacts for the entire Cys2His2 family were disclosed and recognition rules were formulated in our earlier work (Polozov, Sivozhelezov, Chirgadze, & Ivanov, 2015). It was shown that particular contacts of His125 with operator DNA are invariant for the family with high frequency of occurrence and the conservative set of such contacts was considered as an invariant recognition rule.

The binding of the transcription factor to a specific sequence of operator DNA controls the gene transcription. More than 3% of human genes are known to encode the Zn-transcription factors that reflect their biological importance in the cellular functioning. And the factors of Zn-finger type are among most wide-spread. These factors perform various functions and include DNA recognition, RNA packaging, transcriptional activation, regulation of apoptosis, protein folding and assembly, and lipid binding (Desjarlais & Berg, 1992; Krishna, Majumdar, & Grishin, 2003; Laity, Lee, & Wright, 2001; Wolfe, Nekludova, & Pabo, 2000). Examples of analysis for definite Zn-finger structures in connection with their functions will be presented below.

An initial key stage in the protein–DNA interaction is a protein–DNA recognition, 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

*Corresponding author. Email: chir@vega.protres.ru

DNA. At short inter-atomic distances, a number of specific contacts are formed between protein and DNA. The distinctive features of the protein–DNA contacts are displayed when the complex is completed. Such protein–DNA contacts were considered in detail for the factors of the homeodomain family (Chirgadze, Zheltukhin, Polozov, Sivozhelezov, & Ivanov, 2009) and for the Cys2His2 family (Polozov et al., 2015).

The basic Zn-finger unit in the Cys2His2 transcription factor family contains a Zn complex in which a Zn atom is coordinated by two cysteines and two histidines. In this family, the protein factor binds the double-stranded DNA (dsDNA) by the residues of the recognizing α-helix. Usually, the zinc transcription factor proteins contain multiple zinc finger modules, which form a tandem super structure along a DNA chain in the major groove. Each finger of about 30 amino acid residues forms a conservative ββα motif. The most specific feature of such factors is that they include a few Zn-finger units.

The Zn-finger factors are very wide spread and more than 360 thousands of sequences are presented in the GenBank (Benson et al., 2013). All Zn-fingers are classified as different families in the Pfam database (Finn et al., 2016) and they are presented in Table 1. Although all Zn-finger groups are of great interest, one can see that the amount of known sequences and structures of transcription factors with a complex of type Zn-Cys2His2 are an order of magnitude higher than any other types. In fact, there are about 340 thousands of sequences and only 20 thousands of all others. Now we present a short overview of the structural and functional role of the Zn complex of different Zn-finger families in order to disclose their role in DNA binding.

The first class is PF09329, zf-primase, complex Zn-Cys3His1. We consider an example of such Zn-finger which is a part of the multi-domain protein DNA replication complex Mcm10 for assembly and progression of the replication fork. This is an essential eukaryotic DNA replication protein complex which has initiator primase activity. It is presented by the crystal structure of the complex of the internal domain of Mcm 10 ID with single-stranded ssDNA from *Xenopus laevis* (Warren, Huang, Fanning, Chazin, & Eichman, 2009; pdb 3h15). The structure is shown in Figure 1(A). It includes the OB domain, the oligonucleotide/oligosaccharide-binding fold, and the Zn-finger domain. The Zn-finger binds ssDNA by lysines 385 and 386. However, the Zn complex has not participated in this binding mode (Warren et al., 2008).

The second class is PF03119, Zn-finger as a part of the DNA ligase, complex Zn-Cys4. We consider an example of such a Zn-finger which is related to the DNA repair. It is presented by the crystal structure NAD^+-dependent DNA ligase (ligA) from E. coli (Nandakumar, Nair, & Shuman, 2007; pdb 2owo). The large multi-domain protein complex consists of one polypeptide chain with the length of 586 residues. It includes the N-terminal initiation domain Ia, NTase domain (nucleotidyl-transferase), the OB-fold domain, the Zn-finger domain, and the HhH domain (in fact, a complex of two HhH domains). The whole complex incorporates also double-stranded *ds*DNA surrounded by the ligA domains, as shown in Figure 1(B). The Zn-finger does not contact with DNA, its structural role is a linker between the OB-fold and HhH domains.

The third class is PF08996, Zn-finger as a part of the DNA polymerase α complex, complex Zn-Cys4. Here the Zn-finger is related to the replication process of DNA. We consider the complex of human polymerase α, the C-terminal domain of the catalytic subunit, p180c, with B-subunit, p70 (Suwa et al., 2015). The structure of the Zn-finger domain of p180c is shown in Figure 2(A). It contains two Zn complexes; they are divided by two long α-helices of 18 residues each. It was shown that the complex Zn2 is involved in interactions with the B subunit of polymerase α.This interaction between the catalytic and B subunits plays a critical role for the function of DNA polymerases of the B family. This is also confirmed by the fact that the substitution of two cysteines by alanines in the complex Zn2 resulted in disruption of the mouse polymerase α complex (Mizuno, Yamagishi, Miyazawa, & Hanaoka, 1999). However, the Zn complexes are not involved in the binding with DNA.

The fourth class is PF01258, the Zn-finger of the transcription regulation protein family of dksA/traA,

Table 1. Statistics of Zn-finger domains from the Pfam database.

Pfam group of accession	ID of factor classes	Description of different classes and types of Zn complex	Factor sequences	Structures from PDB
PF09329	zf-primase	Primase, *Zn-C3H1*	714	4
PF03119	zf-DNA ligase	NAD^+-dependent DNA ligase, *Zn-C4*	3365	2
PF08996	zf-DNA Pol	DNA polymerase α, *Zn-C4*	748	13
PF01258	zf-dskA/traR	Prokaryotic dksA/traR, *Zn-C4*	4575	15
PF12874	zf-metazoans	Plants to metazoans, *Zn-C2H2*	14,429	6
PF00096	zf-transcription	Transcription factors, *Zn-C2H2*	340,711	412

Designation: *C* – Cys, *H* – His.

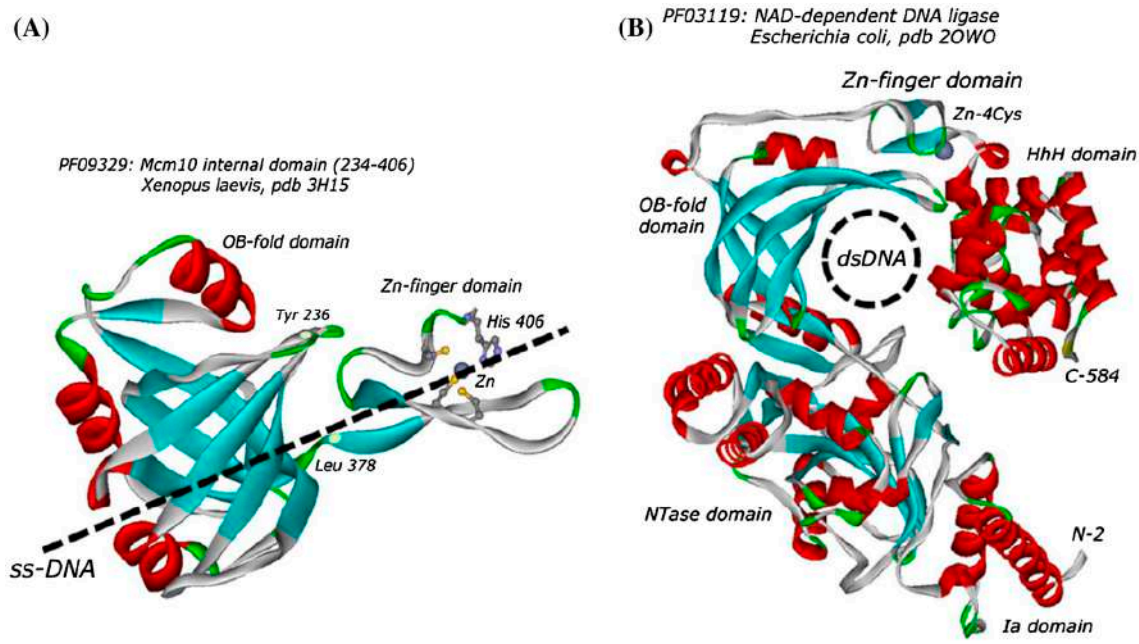

Figure 1. Two protein complexes which belong to two Zn-finger Pfam classes in Table 1. *A*: DNA replication protein complex Mcm10, internal domain, from *Xenopus laevis* for assembly and progression of the replication fork (pdb 3h15). *B*: NAD^{+}-dependent DNA ligase (ligA) from *E. coli* related to the DNA repair (pdb 2owo).

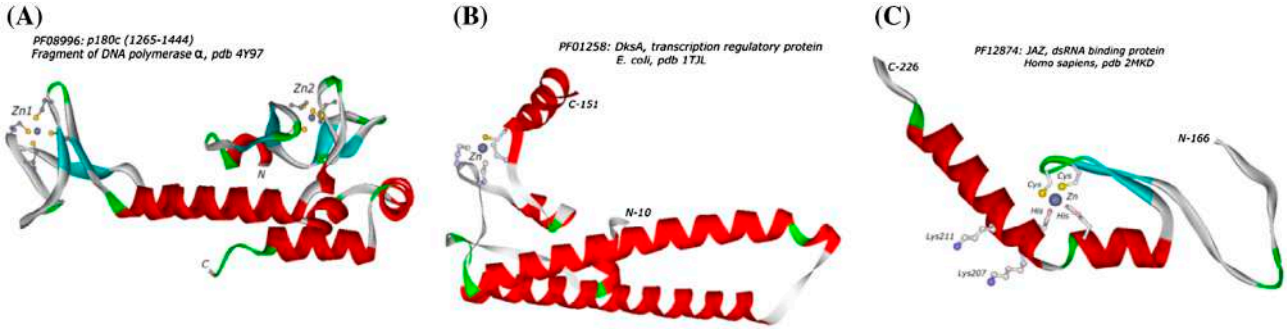

Figure 2. Three protein Zn-fingers from three Pfam classes listed in Table 1. *A*: Fragment p180c of DNA polymerase α from *Homo sapiens*, the C-terminal domain of the catalytic subunit related to the replication process of DNA (pdb 4y97). *B*: Protein DksA from *E. coli*, ppGpp-dependent regulation of bacterial transcription (pdb 1tjl). *C*: Protein JAZ from *Homo sapiens*, the dsRNA-binding Zn-finger protein (pdb 2mkd).

complex Zn-Cys4. These proteins exercise regulatory control over transcription process. We consider protein DksA which is a crucial component of ppGpp-dependent regulation of bacterial transcription (Perederina et al., 2004; pdb 1tjl). The structure is shown in Figure 2(B). It consists of two α-helices and one Zn-Cys4 complex. It enhances regulatory signal through binding to RNA polymerase.

The fifth class is PF12874, dsRNA-binding Zn-finger proteins, complex Zn-Cys2His2. These proteins are involved in numerous biological processes, including cellular localization, antiviral responses, RNA interference (RNAi), and viral inhibition of RNAi. As an example, we consider human JAZ protein (Burge, Martinez-Yamout, Dyson, & Wright, 2014; pdb 2mkd). The structure of fragment is shown in Figure 2(C). It consists of two α-helices, two β-strands and one Zn-Cys2His2 complex. It was shown that the zinc complex is not bound to RNA but three positive lysine residues, K207, K208, and K211, are bound to phosphate groups of the dsRNA backbone (Burge et al., 2014). The Zn complex residues are not bound to dsRNA.

The sixth class is PF00096, transcription factors, complex Zn-Cys2His2. Zinc fingers of this class

recognize a wide variety of different DNA sequences and are one of the most abundant DNA-binding motifs found in eukaryotes. As a classic example, we consider the high-resolution structure of mouse transcription factor Zif268 from *Mus musculus* (Elrod-Erickson et al., 1996; pdb 1aay). The structure of complex of Zif 268 with a fragment of operator dsDNA is shown in Figure 3. Earlier, we have analyzed 22 factor-DNA complexes and have found conservative factor contact His7α with the phosphate group of the DNA backbone (Polozov et al., 2015). Remarkably, this His is also bound with the Zn-atom and forms a complex. Thus, here we can see an example of indirect participation of the Zn complex in the protein–dsDNA binding.

Now we can imagine a variety of functions in different classes containing Zn-finger modules (Table 2). In all classes, the Zn-finger unit consists of a very short polypeptide chain of 30–36 amino acid residues. This suggests that the presence of the Zn complex is necessary for stability of the structure. It is interesting to note that the topological types of these structures for various classes are different, as seen from Table 2 and Figures 1–3. The exceptions are two classes of types ZnCys2His2 which have similar topological types of the chains.

From the above consideration we can make *an important conclusion*: in factors of most classes, the Zn complexes, namely any of four ligand residues, do not participate in binding contacts with functionally important subunits of the multi-domain complexes. As we have seen in these cases, Zn-fingers perform only the 'structural function.' For example, in the second class the Zn-finger domain was used as a connector between two other domains of the multi-domain protein ligase A for DNA repair, and it results in necessary clasping of the dsDNA molecule (Figure 1B). There is only one exception, namely the sixth class: at present we know that one histidine ligand of the Zn complex is bound directly with dsDNA (Table 2). Further, we consider electrostatic features of the Zn complex for this factor class in order to reveal in what way the Zn complex affects the binding with dsDNA.

The modular principle of Zn-finger transcription factors was discovered thirty years ago by studying the transcription factor TFIIIA of the frog oocytes from *Xenopus laevis.* This factor works as an activator and regulator of the gene of the ribosomal 5S RNA (Miller et al., 1985). It consists of nine sequential Zn-finger units specifically bound to double-chained DNA of about fifty nucleotide pairs long. The first atomic crystal structure was determined for a complex from factor Zif268, a mouse immediate early protein, and a consensus DNA-binding site at 2.1 Å resolution about twenty years ago (Pavletich & Pabo, 1991). Later, the structure of this complex was obtained at a much higher resolution of 1.6 Å (Elrod-Erickson et al., 1996).

It is known that electrostatic potential and the shape of molecular surfaces of proteins are two preferable features which play decisive role in protein–DNA binding (Tsuchiya, Kinoshita, & Nakamura, 2004). Recently, Hamed and Arya (2016) have studied the energy of the binding of various Zn-finger motifs of factor TFIIIA of the Cys2His2 type in complex with DNA using molecular dynamics simulation and free energy calculations. It was shown that the loss of total energy of binding is originated in the loss of electrostatic energies on the mutations. This result suggests that electrostatic affinity of this type of transcription factors to operator DNA is very important. The electrostatic potential of the dsDNA molecule is electronegative due to the negative

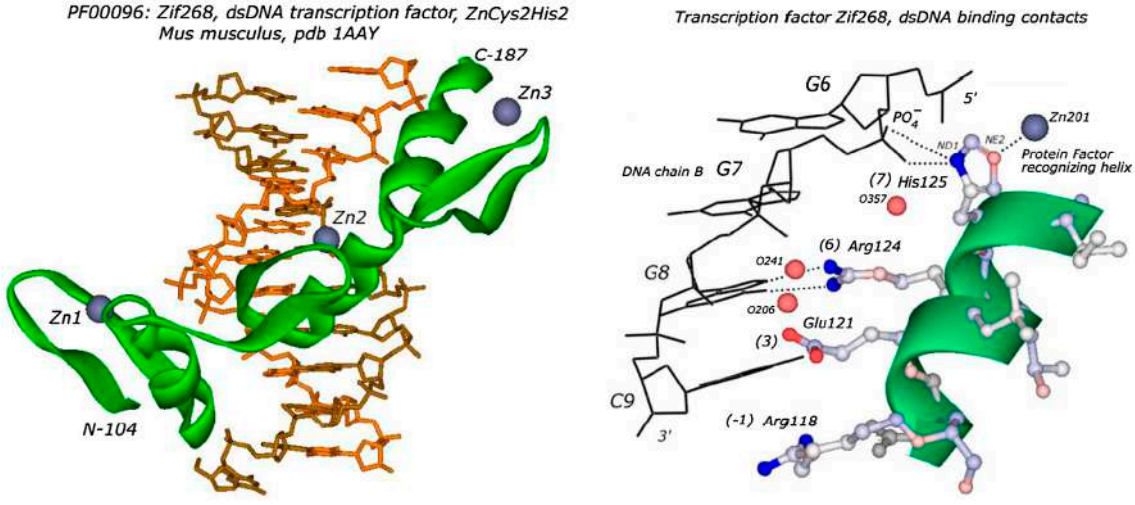

Figure 3. Spatial structure of transcription factor Zif268 from *Mus musculus* bound with operator dsDNA (pdb 1aay). Zn finger contacts with the major groove of double-stranded fragment of operator DNA are shown on the right side of Figure.

Table 2. Description of Zn-finger structure and function.

Pfam accession	ID of factor class	Structure description	Function of representative Zn-fingers in different Pfam classes
PF09329	zf-primase	*βl βl ll*	Binding with ssDNA with two positive lysine residues
PF03119	DNA ligase	*lβ lβ lα*	Fixing positions of OB- and HhH-domains in the complex of NAD^+-dependent DNA ligase
PF08996	zf-DNA Pol	*βl βl βl βl*	Fixing positions of catalytic and B-subunits in the human polymerase *α*
PF01258	zf-dskA/traR	*lα lα*	Crucial component of ppGpp-dependent regulation of bacterial transcription
PF12874	zf-metazoans	*βl βl αl α*	Binding with dsRNA, but probably without Zn complex
PF00096	zf-transcriptions	*βl βl αl*	Binding with dsDNA due to positive charge potential of His125 induced by Zn complex

Designation: *α* – *α*-helix, *β* – *β*-strand, *l* – loop of the main chain.

phosphate backbone. Correspondingly, the binding of transcription factors must have a positive patch at their surfaces. However, it is unknown so far which amino acid side groups, or maybe a Zn-atom, are really responsible for this. Here we have paid attention to the electrostatic contribution of the Zn complex of factor Zif268 at binding to DNA by calculation of the electrostatic potential and revealing the positive region around the conservative residues His125 and Arg124 on the factor surface.

2. Methods

2.1. Calculation of electrostatic potential

In this study, we have considered Zn-fingers as space-filling atomic models. Here each atom is characterized with coordinates of the atom center, radius of van der Waals, and partial charge. In most cases at calculation of the electrostatic potential, the system with point charges is considered. However, upon consideration of the Zn complex we have to pay special attention to accurate charge parameters obtained by using a quantum mechanical approach.

We have used a standard set of force field parameters AMBER 9 in molecular mechanics calculations: program set HyperChem, Version 8.09 (HyperChem, 1996; Weiner, Kollman, Nguyen, & Case, 1986). For calculating the optimized model of Zn complex the quantum mechanics *ab initio* has been applied. The system was composed of Gaussian wave functions, which were used for description of the electron orbital for each atom. After digital solution of Schrodinger equation, the electron distribution of the system has been obtained. Then effective charges have been shifted with special procedure to place them in the centers of atoms in multipole charge model. Finally, using this set of charge parameters we have calculated the electrostatic energy $E_{el\text{-}st}$ for the Zn finger by the known formula:

$$E_{el-st} = \sum q_i q_j / \varepsilon R_{ij}$$

Here the electrostatic interactions are calculated between two atoms i and j with the point charges q_i and q_j. The magnitude of the electrostatic energy varies inversely with the distance between the atoms, Rij. The effective dielectric constant is ε. *In vacuo* approximation value ε is equal to 1. We have used *ε* as 'coordinate-dependent' with factor 1, roughly accounting the nearest local water medium. In our case, the dielectric constant *ε*(*r*) was assumed to be 2 inside the protein medium and 80 outside it.

The classical electrostatic potential for the system of point charges is the potential energy at position R according to the formula:

$$V = \sum \frac{q_i}{|R - r_i|}$$

In order to take into account more accurate influence of water medium on electrostatic potentials we have solved the Poisson–Boltzmann equation. Below we consider in details the calculation approaches for the Zn complex model, the Zn-finger protein unit and the whole factor molecule of Zif268.

2.2. Quantum chemical calculations: optimized model for Zn-Cys2His2 complex

The *ab initio* calculations of optimal coordination geometry and electron density distribution in the Zn-Cys2His2 complex have been performed using the Gaussian-09 package of quantum chemical programs (Frisch et al., 2009). The chemical structure of the particular model complex is shown in Figure 4. It was based on a number of N-methylimidazol thiolato Zn structures found in the Cambridge Structural Database (Groom, Bruno, Lightfoot, & Ward, 2016).

A few series of calculations were performed in parallel to get an idea on how theoretical approximations specific to a particular functional and/or basic set of Gaussian functions influence the optimized structure parameters and electrostatic properties of the complex. These

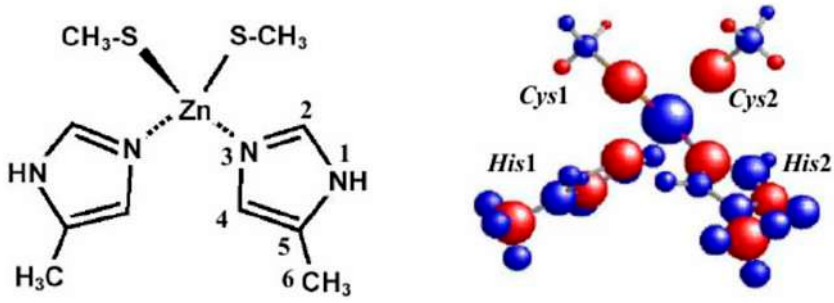

Figure 4. Model of the Zn-Cys2His2 complex in quantum chemical calculations. *Right* – Chemical diagram of the Zn complex model. *Left* – View of the 'best atomic charge' (BAC) model. Negative and positive charges are rendered as red and blue spheres, respectively; their volumes are proportional to charge magnitudes. All atomic net point charges are given in Table 4.

series involved (1) the conventional Hartry–Fock MO LCAO SCF method with 6–31G(*d*, *p*), 6–31++G(*d*, *p*) and sets of basic functions, and (2) the B3LYP density functional method with 6–31G(*d*, *p*) and LANL2DZ basis sets, of which B3LYP/LANL2DZ approach is cited as *de facto standard* in quantum chemistry (Tsipis, 2014).

In geometry optimizations, the total energy of electrons and nuclei of the complex was minimized with respect to the coordinates of all atoms of the molecular complex. Regardless of the method/basis used, the final geometries showed a tendency to be somewhat greater than the coordination distance Zn–N, by 0.05–0.10 Å, with respect to experimental distances in complexes of ten N-methylimidazol thiolato Zn structures, on average 2.04 Å. The bond angle S–Zn–S is predicted by quantum mechanical approach up to 20 degrees greater than from experimental sources, on average 118 degrees.

To this end, additional optimizations have been performed in which the coordination geometry and overall conformation parameters of the four ligands were fixed upon structure optimization, whereas the remaining geometric parameters, such as bond lengths and bond angles within the ligands were allowed to vary to relax the structure as close to the true equilibrium state.

2.3. Molecular electrostatic potential: derivation of site charge and multipole charge models from the electrostatic potential of Zn complex

The optimized geometric model of the complex Zn-Cys2His2 was used to generate the distribution of the electrostatic potential $V(x_i,y_i,z_i)$ for a grid of probe points x_i,y_i,z_i (i = 1, 2, …, N_p) using the increment 0.25 Å in each dimension. The points occurring closer to any atom k than the sum of its van der Waals radius plus 1.02 Å were excluded from the grid. The outer boundary of the grid space was determined by the condition $|r_i - r_k| < 7.5$ Å, where k is any atom occurring most closely to grid point i. Thus, the total number of selected grid points was about 70 thousands.

The list of {*x, y, z, V*} quantities from Gaussian-09 was treated with program FitMEP (Dzyabchenko, 2008) designed for approximation of the three-dimensional electrostatic potential distribution with the analytical potential of a set of point charges or multipoles up to quadrupole depending on the model specified. In this procedure, the minimized function F, calculated as the mean-square deviation of the model potential $V_{i,\ \mathrm{model}}$ from the quantum mechanical potential V_i,

$$F = N_p^{-1} \sum \left(V_i - V_{i,\ \mathrm{model}}\right)^2$$

This was minimized with respect to the parameters of the model.

Table 3 presents characteristics of three charge models, BAC, BSC, and CDQ, which have been derived from the molecular electrostatic potential calculated with the HF/6–31G(d, p) approach (Sigfridsson & Ryde, 1998). The forth Mulliken charge model of regular population analysis of the MO LCAO SCF method is presented as well, demonstrating the worst agreement with the 'quantum mechanical' (QM) potential as compared to the other three. By letting charge magnitudes to optimize in minimization of F, we come to a much more accurate 'best atomic charge' (BAC) model where the charges stay fixed at atomic nuclei (Figure 4). With all charge positions allowed to vary in three dimensions, we get further significant progress in the approximation accuracy for model 'best site charges' (BSC) with R_r less than 1%, quite as good for calculation of the intermolecular energy. The dipole moment (D) and quadrupole moment components Q_{ab} of the molecular complex, calculated with model charges, demonstrate excellent agreement with corresponding properties calculated by integration of the wave function in Gaussian-09 (model QM). Further progress with r.m.s. deviations three point times smaller is reached with the atomic multipole mode 'charge + dipole + quadrupole' (CDQ), a kind of the 'distributed multipole analysis' (DMA) model (Stone & Alderton, 1985) in which each atom is assigned to

Table 3. Different charge models of Zn-Cys2His2 complex derived from the molecular electrostatic potential.

Model	R, kcal/mol	R_r, %	D, Debye	Q_{xx}, Debye Å	Q_{zz}, Debye Å
Mulliken	3.978	30.7	12.88	29.70	−12.55
BAC	0.510	4.04	12.40	38.41	−18.40
BSC	0.088	0.70	12.51	38.11	−18.40
CDQ	0.026	0.20	12.51	38.05	−18.41
QM	0	0	12.50	38.01	−18.40

Models: Mulliken – Atomic charges from Mulliken population analysis; BAC – 'Best Atomic Charges' (charge sites fixed on nuclei); BSC – 'Best Site Charges' (charge-site positions and charge magnitudes optimized); CDQ – 'Charge + Dipole + Quadrupole' (point multipole up to quadrupole at each atom); QM – Quantum Mechanical (HF/6-31G(d, p)) electric charge moments.
Designations: $R = (F)^{1/2}$ is the r.m.s. deviation of the model potential from the quantum mechanical QM electrostatic potential; $R_r = RN_p^{1/2}(\sum V_i^2)^{-1/2}$ is the relative r.m.s. deviation. D is the dipole moment of molecule; Q_{xx} and Q_{zz} are the diagonal-matrix components of the traceless ($Q_{xx} + Q_{xx} + Q_{zz} = 0$) quadrupole moment of molecule.

Table 4. Distribution of net atomic charges in the 'Best Atomic Charges' model where charge sites are fixed on the nuclei.

Atom	Charge, e	Charge, e
Zn	1.255	–
Cys	*Cys1*	*Cys2*
S	−0.795	−0.836
C	0.128	0.191
H	0.050	0.037
H'	−0.008	−0.036
H"	−0.046	−0.050
His	*His*1	*His*2
N3	−0.700	−0.879
C2	0.378	0.506
H2	0.109	0.040
N1	−0.547	−0.505
H1	0.413	0.409
C5	0.416	0.330
C4	0.034	0.169
H4	0.114	0.100
C6	−0.848	−0.929
H6	0.241	0.275
H6'	0.217	0.267
H6"	0.251	0.250

a point charge, a point dipole and a point quadrupole. The problem is, however, that energy calculations with DMA models of large systems like biomolecules and crystals are quite sophisticated and computationally time-expensive. Consequently, the BAC or BSC models seem to be a reasonable compromise with respect to both sufficient accuracy in the estimation of electrostatic energy and its computational costs.

For our further calculations of the electrostatic potential we have used the 'Best Atomic Charges' (BAC) model shown in Figure 4. Partial charges for this model are presented in Table 4. It should be noted that Zn ion has charge + 1.26 electrons.

2.4. *Zn-finger electrostatic potential by solving the Poisson–Boltzmann equation*

In order to take into account a more accurate influence of water medium on the electrostatic potentials we have solved the Poisson–Boltzmann equation. We have used for this step the corrected charge distribution taken from either AMBER force field or quantum mechanics *ab initio* approach for Zn-complex. This equation has been solved as specified in (Polozov, Montrel, Ivanov, Melnikov, & Sivozhelezov, 2006):

$$-\nabla(\varepsilon(r)\nabla\phi(r)) = 4\pi(\rho_0(r) + \rho_1(\phi(r))$$

where $\rho_0(r) = \Sigma_i z_i q \delta(r_i)$ is the charge distribution of the protein molecule, $\rho_1(\phi(r)) = \Sigma_j z_j n_j q \exp(-z_j q\phi(r)/kBT)$ is the distribution of the mobile electrolyte charges, $\phi(r)$ is the electrostatic potential, $\varepsilon(r)$ is the dielectric constant to be 2 inside the protein and 80 outside it, r(x, y, z) is the radius vector of each observation point, z_i and r_i are the charge and radius vectors of the i-th atom of the molecule, z_j and n_j are charges and concentrations of electrolyte ions, q is the proton charge, k_B is the Boltzmann constant, T is the absolute temperature.

Charges z_i were taken from the AMBER force field. The electrolyte is assumed to be univalent ($z_1 = -1$; $z_2 = 1$) at physiological concentration of 150 mM. Solution is sought with the finite-difference multi-grid method using a sequence of nested finite difference grids, so that the interval between grid points is 0.4Å. Analytical solution, available when ε = const, has been used as the first approximation as well as to establish the boundary conditions. Solving the Poisson–Boltzmann equation has been used for electrostatic calculation of Zn-finger. For the whole Zif268 factor we have used Molecular Mechanics Force Field, AMBER 9 package with parameters: Dielectric constant: coordinate-dependent; Scale factors: dielectric constant 1, electrostatic contribution 0.83, Van der Waals contribution 0.50; atomic charges are the standard with exception for Zn complex, for which they are presented in Table 4.

Examples of the calculations of the whole molecule of mouse Zif268 by the two methods are presented in Figure 5. One picture has been obtained in this work, and the second – produced by the molecular graphic program PyMOL (PayMOL molecular graphics systems website, originally based on DeLano, Ultsch, de Vos,

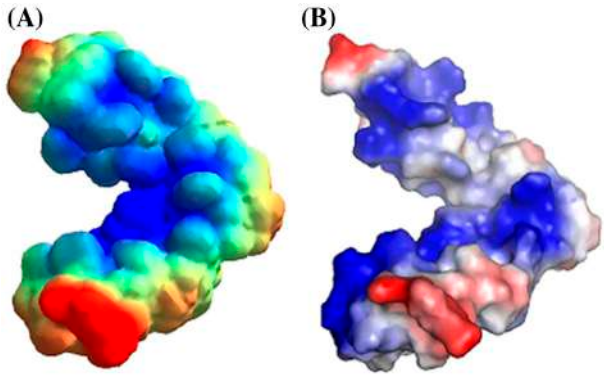

Figure 5. Electrostatic potential of mouse transcription factor Zif268. *A*: Calculated with force field AMBER 9, semi-empirical method, system of point charges, see text. The potential is drawn on the surface removed at 1Å from the van der Waals surface. *B*: The potential is calculated with force field AMBER 9 by program PyMol. The surface potential is drawn at 0.7Å from the van der Waals surface. In both cases, the electrostatic potentials have shown by the same color.

and Wells (2000); http://pymol.org/). One can see similarity of the two results that confirms reliability of our approach. However, some differences are observed, which could be explained by different presentations of the electrostatic potential surface (different distance from the van der Waals surface) and possibly more accurate accounting of the dielectric constant value in the current work. It should be also noted that a more correct approach taking into account water medium required the solving the Poisson–Boltzmann equation.

3. Results and discussion

3.1. Stereochemistry and structure of Zn complexes

Atom Zn is an element in the fourth period of the periodic table. The group includes 10 elements beginning from 21Sc to 30Zn. In this period, the filling of $3d^{10}$-electronic orbital occurs and spatial configuration of Zn corresponds to that one of this orbital.

The orbital, which accept side groups of Cys and His amino acid residues as ligands are directed along the spatial bisectors in the Zn complex coordinate system. That can be observed in all spatial structures of different types of Zn complexes as shown in Figure 6.

3.2. Electrostatic potential of Zn complex

Electrostatic potential for complex Zn-Cys2His2 has been calculated with the atomic coordinates of Zif268 (pdb 1aay). We have considered a complex from the first zinc finger, which is one of three similar units in the whole factor. The electrostatic potential for the model of the Zn complex is presented in Figure 7(A). It should be noted that potential values correspond to the spatial points distributed on the sphere with the radius of 5.5 Å and the center at the Zn-ion. Strong negative potentials have arisen from cysteine residues, and strong positive potentials from atomic groups of histidine residues. We can see that magnitudes of this potential are essentially induced by the zinc ion. This phenomenon is well understood if we have in mind that the electrostatic potential is taken to be at 5.5 Å from the center of Zn and integrated from the atomic charges of ligands in the central area of the Zn complex (Figure 4).

The picture drastically changed when the Zn-atom was removed from the model (Figure 7B). In both cases of Cys and His, electrostatic potentials diminish and potentials from different cysteine and histidine residues became oppositely polarized. That means that the Zn atom has a significant inducing effect on the electrostatic potential of atomic groups of ligand residues.

This result is very important for the formation the protein factor–DNA complex because conservative His125 forms a direct contact with the negative phosphate group of DNA (Figure 3). Thus, we can conclude that the Zn complex of the first unit supplies a binding function upon formation of the protein–DNA complex. However it can be noted that Zn ion influences upon protein–DNA binding indirectly.

3.3. Electrostatic potential of Zn-finger unit

Electrostatic potentials of the first Zn-finger unit in different projections and at two different distances from the van der Waals molecular surface are presented in Figure 8. The result at 3.0Å contains more details, while the more distant result at 5.5 Å is theoretically more correct. Also one should keep in mind that the potential is diminishing with the distance of observation, as shown in Section 2.

The positive potential is located at the Zn-unit side which has contacts with the major groove of dsDNA as seen in Figure 9. This region is arisen from the side groups of histidine 125 and arginine 124 residues. The residues are located at the side surface of the ‘so called’ recognition α-helix of the Zn-finger unit (Polozov et al., 2015).

Our result of the electrostatic calculations of Zn-finger 1 is confirmed by the experimental data on binding reaction of Zif268 mutants with the corresponding Arg positions (Elrod-Erickson & Pabo, 1999). An important role is played by Arg118 (position -1t) and Arg124 (position 6 of the α-helix). It should only be noted that electrostatic role of Arg118 is weak, and it binds with some nucleotide group but not the phosphate backbone of DNA.

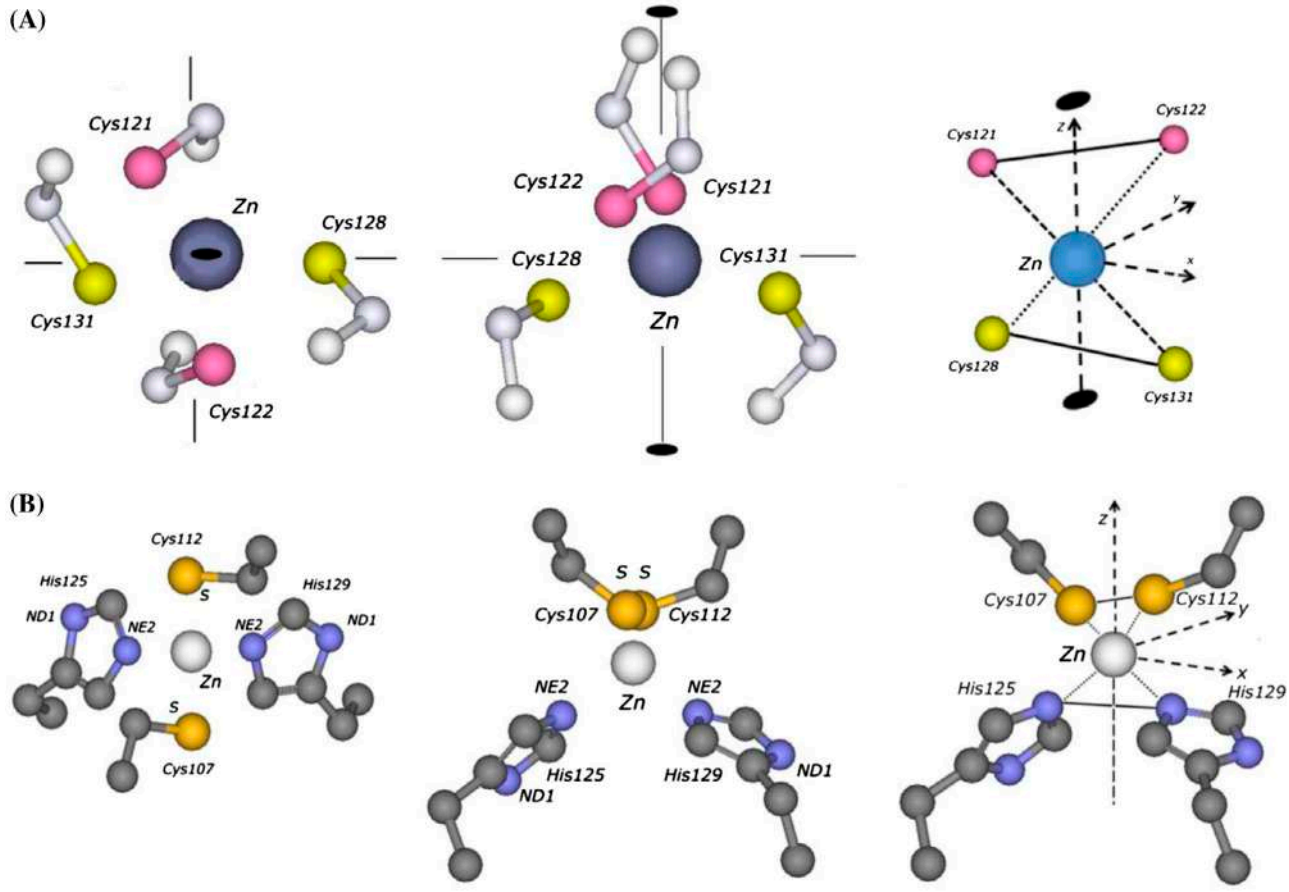

Figure 6. Spatial geometry of Zn complexes. *A*: The Zn-Cys4 complex. It was taken from the structure of polymerase α, the C-terminal domain of the catalytic subunit (pdb 5exr). In order to show the related twofold rotation axis the SH cysteine groups are painted in different colors. *B*: The Zn-Cys2His2 complex. It was taken from Zn-finger 1 of transcription factor Zif268 (pdb 1aay).

His125 forms a charge bond with the negative phosphate group of DNA, which completes the charge-alternating chain as follows:

$$^{-}\mathrm{OPO}^{-} \ldots \mathrm{H}^{+\delta}\mathrm{ND1\ His125\ NE2}^{-\delta} \ldots \mathrm{Zn}^{+\mathrm{Q}} \ldots$$
$$\mathrm{NE2}^{-\delta}\mathrm{His129\ ND1H}^{+}\delta.$$

Here δ is partial charges which are induced on the atoms of imidazol rings of histidines by the negative phosphate group and positive Zn ion. Charge Q of Zn-ion is equal to + 1.26 electrons. Distribution of point partial charges in Zn complex can be seen in Table 4 in Methods, Section 2.3. Such an extended sign-alternating charge chain increases inducing the effect of the Zn-ion located in the complex. Having in mind similarities of the sequences and structures of other fingers (Elrod-Erickson et al., 1996) we suggest that all three fingers have similar patterns of electrostatic potentials. In fact, we can see that in the whole factor, which is considered in the next section.

3.4. Electrostatic potential of the whole factor Zif268

This transcription factor is a protein fragment which consists of three Zn finger and contains 90 residues. Having in mind similarities of the sequences and structures of all fingers (Elrod-Erickson et al., 1996), we suggest that all three fingers will have the similar patterns of electrostatic potentials.

There are four residues in each of three Zn finger units, which contribute to the recognition (Table 5). One is the residue His7α which determines the conservative binding contact with the phosphate group of DNA, and three residues at positions -1t (here t means a turned residue), 3α, 6α provide variable recognition binding (Polozov et al., 2015). In the sequences of the whole Zn-Cys2His2 transcription factor family there are five rather conservative positions with high sequence identity of more than 70%. These include residues Phe (−3β), Leu (4α), His (7α), His (11α), and Thr (12α) with very high frequency occurrences for 46 considered Zn fingers (Polozov et al., 2015). Frequency occurring values for

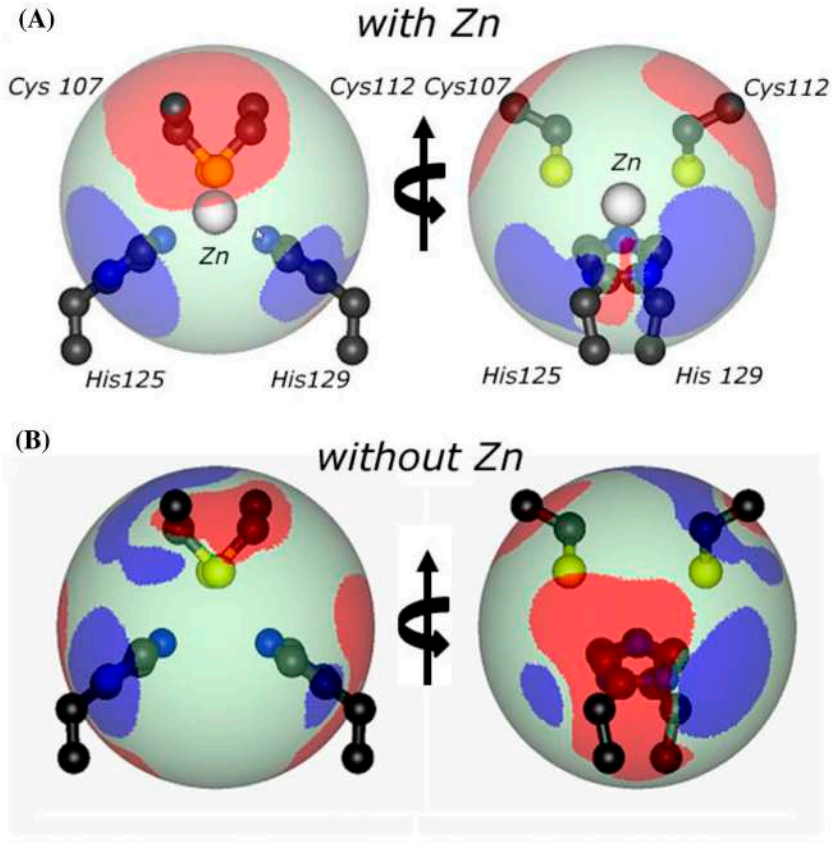

Figure 7. Electrostatic potential for the Zn-Cys2His2 complex is taken from the first Zn finger unit of transcription factor Zif268 (pdb 1aay). *A*: Complex with Zn. *B*: Complex without Zn. Two views are shown with rotation around the vertical axis at 90 degrees. Potentials are presented as color areas on the spherical surface with a radius of 5.5 Å and the center placed at the Zn-ion. Blue color is the positive potential, red – the negative potential, and gray corresponds to the neutral potential.

these residues are 85, 80, 100, 87 and 70%, respectively. Surprisingly, there is only one conservative helix residue His125 (position 7α) with the absolute sequence identity 100%, which is involved in any case in the recognition with the binding DNA phosphate group. Structure-specific variable contacts with nucleotide bases are provided with amino acid positions at −1t, 3α and 6α. The rest of conservative positions Phe (−3β) and Leu (4α) seem to contribute to the stability of the protein Zn finger unit but are believed to have no relation to the DNA recognition or binding. They supply the formation of some kind of a hydrophobic core.

The electrostatic potential of the whole factor Zif268 from *Mus musculus* is shown in Figure 10. It should be noted that we have used here a solely simplified point charge model with standard partial charges placed on the atoms of amino acid residues. The regions of Zn ions can be considered only as qualitative. The picture displays three positive blue patches close to the zinc–histidine regions of each finger units having contacts with double-stranded DNA. Negative red spots on two opposite ends of the factor are related to the side groups of negatively charged residues Glu110, Asp113, and Asp166. The rest surface of the Zn-finger unit, opposite to contact with *ds*DNA, is neutral. Thus, we can conclude that the electrostatic potential of Zn complexes in the considered factor Zif268 plays a significant role in binding with operator dsDNA.

We have seen that all three Zn-fingers have very similar positively charged electrostatic surfaces. It should be noted that amino acids in position 4 are of different types, which does not contribute essentially to the charge potential. Amino acids in this position are Leu in Zn-finger 1, 2 and Arg in Zn-finger 3 (Table 5). However, the experimental data on the stability of the Zif268 mutants (Negi et al., 2011) shows that the Leu and Arg play a similar role in stability of the Zn-finger domain, and this can be explained by similar interaction of the hydrophobic side chain of Leu and the aliphatic carbon side chain of the Arg residue.

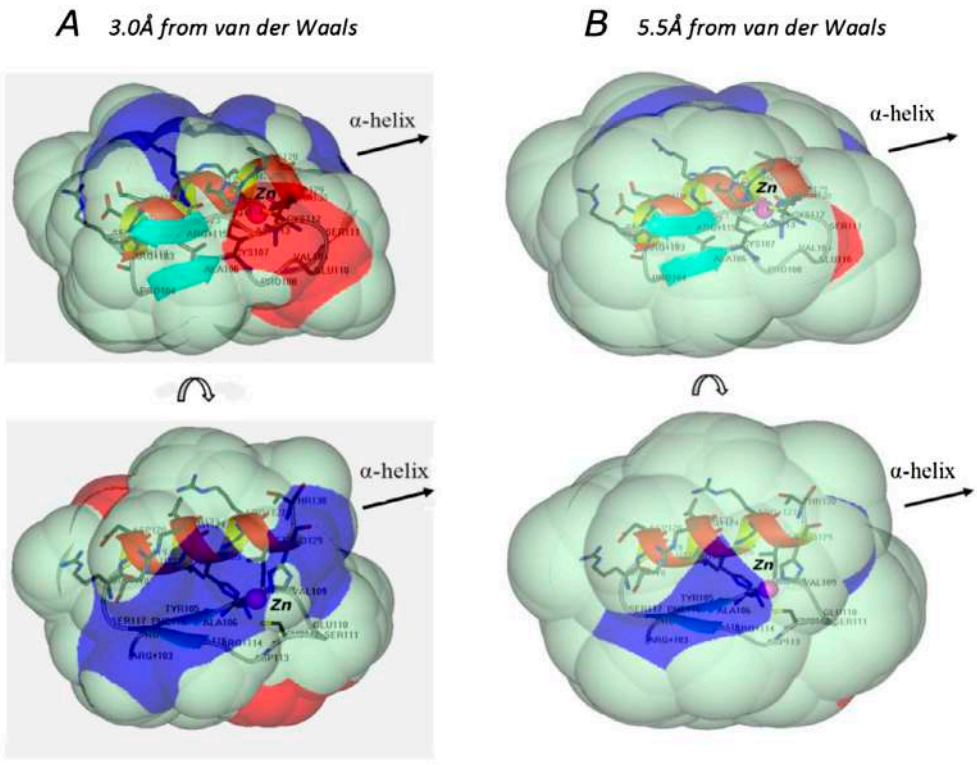

Figure 8. Electrostatic potential calculated for the first Zn-finger of factor Zif268. *A*: Electrostatic potential surface distantly removed from the van der Waals surface at 3.0Å. *B*: Electrostatic potential surface distantly removed from the van der Waals surface at 5.5Å. In both columns: in the lower row models are rotated at 90 degrees around horizontal axes relative to the models of the upper row. Arrows are the axes of recognizing α-helices of Zn-fingers.

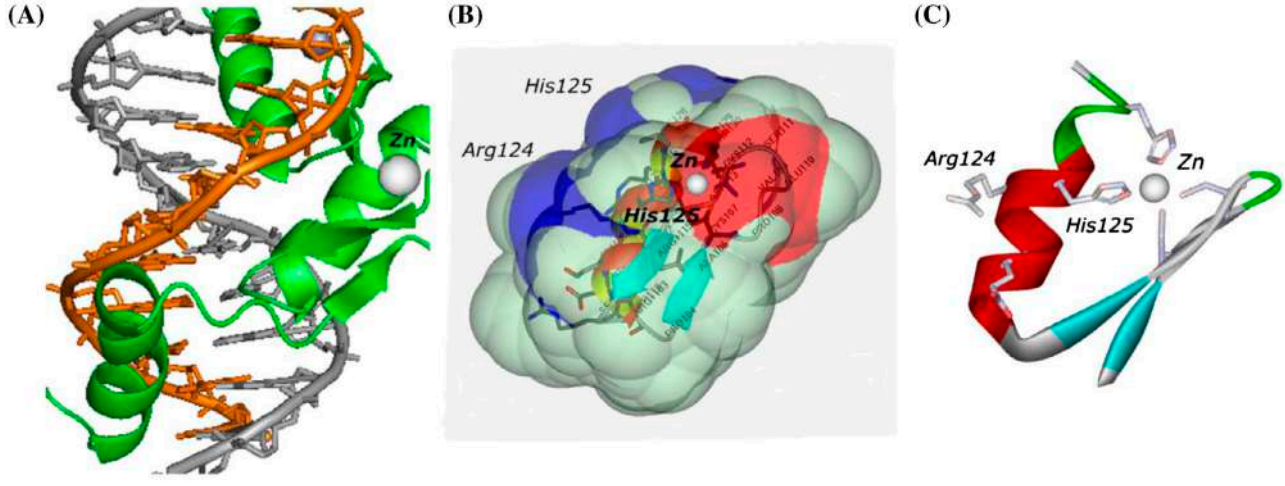

Figure 9. Schematic diagram of the first Zn finger unit of factor Zif268 with a fragment of dsDNA. *A*: ds DNA. *B*: Zn-finger unit with the surface of the electrostatic potential distantly removed from the van der Waals surface at 3.0Å. Color coding: positive values are shown in blue, negative values in red, neutral in gray color. *C*: Skeletal model with the location of the binding residues and the Zn complex.

Table 5. Binding contacts of recognizable triplets of operator DNA and recognizing protein helix in three units Zn fingers of Zif268 transcription factor belonging to the Zn-Cys2His2 family.

Coding DNA chain

```
Zif268  DNA-tripet        Amino acid position - internal system numeration
PDB     (underlined)                             -5-4-3-2-1 1 2 3 4 5 6 7 8 9 10 11 12
code                        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: DNA contact positions of recognizing α-helix are shown in red boxes. Color coding of contacts: yellow – amino acids with phosphates, and cyan – amino acids with bases. Underlined letters show charged residues. Red boxes and asterisks designate preferable binding contacts of protein–DNA found also in the whole Zn-Cys2His2 family.

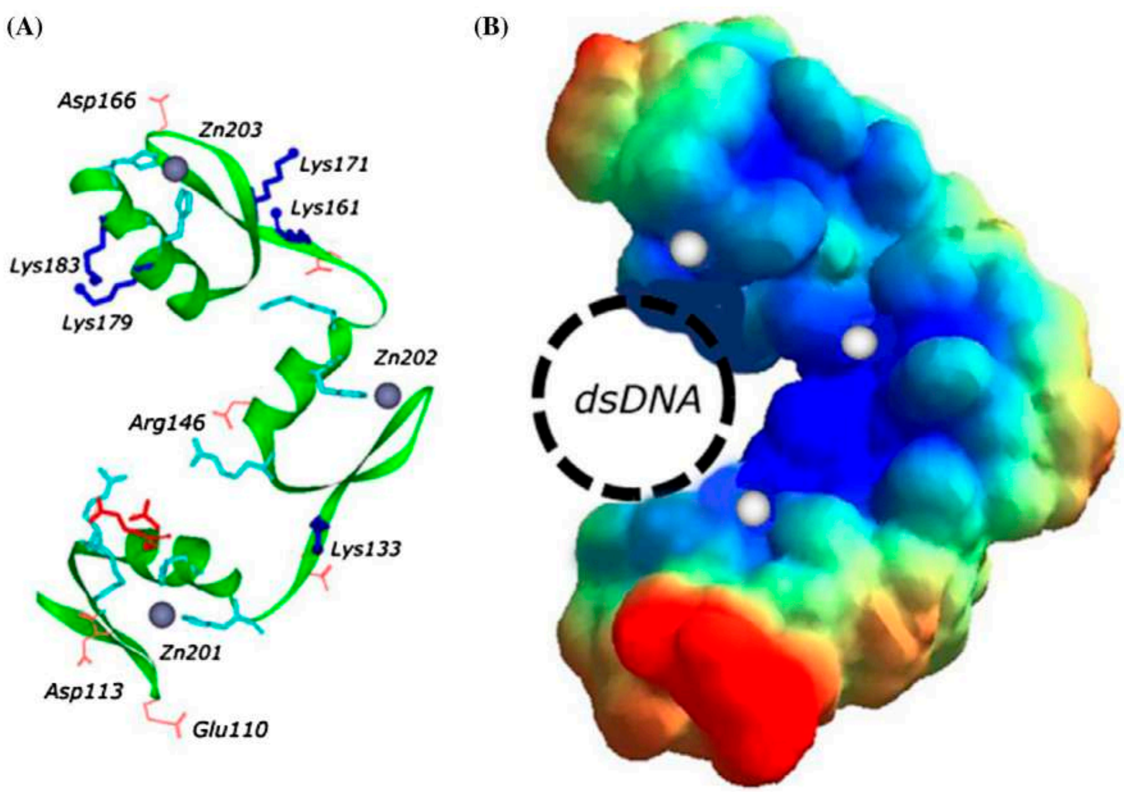

Figure 10. Atomic model and electrostatic potential of the whole transcription factor Zif268 from *Mus musculus*. *A*: Secondary structure with some positive and negative charged residues. *B*: Electrostatic potential surface removed at 1.0Å from the van der Waals surface. The three zinc atoms are shown as gray balls.

4. Conclusion and discussion

We have analyzed the structural and functional roles of the transcription factors of several Zn-fingers belonging to different families. The electrostatic role of the Zn complex in the Zn fingers of factor Zif268 has been considered on the basis of the electrostatic potentials of the Zn complex, the Zn-finger unit, and the whole transcription factor. Two main results have been obtained, as presented below. **First.** Among the six considered classes of Zn-fingers only one class Zn-Cys2His2, the Pfam database PF00096, has the Zn complex, which plays a significant role due to its electrostatic influence. And this is the most spreading family consisting of more than 340 thousands transcription factors. **Second.** The Zn complex

in the transcription factors of family Zn-Cys2His2 plays a significant role in the protein–DNA binding due to its electrostatic influence of the Zn ion on the binding of histidine residue His 125 (Zn-finger 1) with the phosphate DNA backbone. Thus, we can conclude that the electrostatic potential of the Zn-finger unit with the Zn-Cys2His2 complex plays a significant role in the binding and recognition of transcription factors with the operator dsDNA. Because of conservatism of His125, the obtained effect is suitable for all the protein factors from this family. This is particularly related to the variety of human factors and could be of medical interest.

The electrostatic nature of the binding and recognition suggests being the same also for all transcription factors of the Zn-Cys2His2 family. However, it should be noted that the Zn ion has an indirect effect on the protein–DNA binding. Finally, one can suggests that the electrostatic effect could be significant for the transcription factors of the two classes, Pfam database PF12874 and PF00096, because they both have the same complex type. However, this assumption should be verified.

Acknowledgments

The authors would like to thank Prof. Alexey V. Finkelstein and Prof. Vyachaslav A. Kolb, both from Institute of Protein Research, Russian Academy of Sciences, for valuable discussion.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Russian Foundation for Basic Research [grant numbers 17-07-01331a and 14-03-01091a].

References

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research, 41*, 36–42. www.ncbi.nlm.nih.gov/genbank/

Burge, R. G., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2014). Structural characterization of interactions between the double-stranded RNA-binding zinc finger protein JAZ and nucleic acids. *Biochemistry, 53*, 1495–1510.

Chirgadze, Yu N, Zheltukhin, E. I., Polozov, R. V., Sivozhelezov, V. S., & Ivanov, V. V. (2009). Binding regularities in complexes of transcription factors with operator DNA: Homeodomain family. *Journal of Biomolecular Structure & Dynamics, 26*, 687–700.

DeLano, W. L., Ultsch, M. H., de Vos, A. M., & Wells, J. A. (2000). Convergent solutions to binding at a protein–protein interface. *Science, 287*, 1279–1283.

Desjarlais, J. R., & Berg, J. M. (1992). Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proceedings of the National Academy of Sciences, 89*, 7345–7349.

Dzyabchenko, A. V. (2008). Multipole approximation of electrostatic potential of the molecules. *Journal of Physical Chemistry (Russia), 82*, 875–884.

Elrod-Erickson, M., & Pabo, C. O. (1999). Binding studies with mutants of Zif268. Contribution of individual side chains to binding affinity and specificity in the Zif268 zinc finger–DNA complex. *Journal of Biological Chemistry, 274*, 19281–19285.

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

Elrod-Erickson, M., Benson, T. E., & Pabo, C. O. (1998). High-resolution structures of variant Zif268–DNA complexes: Implications for understanding zinc finger–DNA recognition. *Structure, 6*, 451–464.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research, 44*, D279–D285.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., … Fox, D. J. (2009). *Gaussian 09, Revision A.02*. Wallingford, CT: Gaussian, Inc.

Groom, C. R., Bruno, I. J., Lightfoot, M. P., & Ward, S. C. (2016). The Cambridge structural database. *Acta Crystallographica, B72*, 171–179.

Hamed, M. Y., & Arya, G. (2016). Zinc finger protein binding to DNA: An energy perspective using molecular dynamics simulation and free energy calculations on mutants of both zinc finger domains and their specific DNA bases. *Journal of Biomolecular Structure and Dynamics, 34*, 919–934. doi:10.1080/07391102.2015.1068224

Hypercube, Inc. (Ed.). (1996). *HyperChem Computational Chemistry* 'Part 1. Practical Guide; Part 2. Theory and Methods' (pp. 1–350). Ontario: Hypercube.

Krishna, S. S., Majumdar, I., & Grishin, N. V. (2003). Structural classification of zinc fingers: Survey and summary. *Nucleic Acids Research, 31*, 532–550.

Laity, J. H., Lee, B. M., & Wright, P. E. (2001). Zinc finger proteins: New insights into structural and functional diversity. *Current Opinion in Structural Biology, 11*, 39–46.

Miller, J., McLachlan, A. D., & Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus oocytes*. *EMBO Journal, 4*, 1609–1614.

Mizuno, T., Yamagishi, K., Miyazawa, H., & Hanaoka, F. (1999). Molecular architecture of the mouse DNA polymerase α-Primase complex. *Molecular and Cellular Biology, 19*, 7886–7896.

Nandakumar, J., Nair, P. A., & Shuman, S. (2007). Last stop on the road to repair: Structure of E. coli DNA ligase bound to nicked DNA-adenylate. *Molecular Cell, 26*, 257–271.

Negi, S., Imanishi, M., Sasaki, M., Tatsutani, K., Futaki, S., & Sugiura, Y. (2011). An arginine residue instead of a conserved leucine residue in the recognition helix of the finger 3 of Zif268 stabilizes the domain structure and mediates DNA binding. *Biochemistry, 50*, 6266–6272.

Pavletich, N. P., & Pabo, C. O. (1991). Zinc finger–DNA recognition: Crystal structure of a Zif268–DNA complex at 2.1 A. *Science, 252*, 809–817.

Perederina, A., Svetlov, V., Vassylyeva, M. N., Tahirov, T. H., Yokoyama, S., Artsimovitch, I., & Vassylyev, D. G. (2004). Regulation through the secondary channel – Structural framework for ppGpp-DksA synergism during transcription. *Cell, 118*, 297–309.

Polozov, R. V., Montrel, M., Ivanov, V. V., Melnikov, Y., & Sivozhelezov, V. S. (2006). Transfer RNAs: Electrostatic patterns and an early stage of recognition by synthetases and elongation factor EF-Tu. *Biochemistry, 45*, 4481–4490.

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

Sigfridsson, E., & Ryde, U. (1998). Comparison of methods for deriving atomic charges from the electrostatic potential and moments. *Journal of Computational Chemistry, 19*, 377–395.

Stone, A. J., & Alderton, M. (1985). Distributed multipole analysis. *Methods and applications. Molecular Physics, 56*, 1047–1064.

Suwa, Y., Gu, J., Baranovskiy, A. G., Babayeva, N. D., Pavlov, Y. I., & Tahirov, T. H. (2015). Crystal structure of the human Pol α B subunit in complex with the C-terminal domain of the catalytic subunit. *Journal of Biological Chemistry, 290*, 14328–14337.

Tsipis, A. C. (2014). DFT flavor of coordination chemistry. *Coordination Chemistry Review, 272*, 1–29.

Tsuchiya, Y., Kinoshita, K., & Nakamura, H. (2004). Structure-based prediction of DNA-binding sites on proteins using the empirical preference of electrostatic potential and the shape of molecular surfaces. *Proteins: Structure, Function, and Bioinformatics, 55*, 885–894.

Warren, E. M., Vaithiyalingam, S., Haworth, J., Greer, B., Bielinsky, A. K., Chazin, W. J., & Eichman, B. F. (2008). Structural basis for DNA binding by replication initiator Mcm10. *Structure, 16*, 1892–1901.

Warren, E. M., Huang, H., Fanning, E., Chazin, W. J., & Eichman, B. F. (2009). Physical interactions between Mcm10, DNA, and DNA Polymerase α. *Journal of Biological Chemistry, 284*, 24662–24672.

Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986). An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry, 7*, 230–252.

Wolfe, S. A., Nekludova, L., & Pabo, C. O. (2000). DNA recognition by Cys2His2 zinc finger proteins. *Annual Review of Biophysics and Biomolecular Structure, 29*, 183–212.