Journal of Molecular Graphics and Modelling 29 (2011) 928–934

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Theoretical design of a specific DNA–Zinc-finger protein interaction with semi-empirical quantum chemical methods

Gábor Nagy[a,b], Béla Gyurcsik[b], Eufrozina A. Hoffmann[a], Tamás Körtvélyesi[a,*]

[a] Department of Physical Chemistry and Material Sciences, University of Szeged, Aradi Vértanuk tere 1, H-6720 Szeged, Hungary
[b] Department of Inorganic and Analytical Chemistry, University of Szeged, Dóm tér 7, P.O. Box 440, H-6701 Szeged, Hungary

ARTICLE INFO

Article history:
Received 15 October 2010
Received in revised form 24 February 2011
Accepted 3 March 2011
Available online 10 March 2011

Keywords:
Zinc-finger protein
DNA
Docking
Semi-empirical quantum chemical methods
PM6-DH2

ABSTRACT

The interactions of a zinc-finger (ZF) protein with DNA containing the specific recognition site of the ZF and with a non-specific DNA were studied with the semi-empirical quantum chemical method of PM6/Mozyme. The ZF protein (1MEY)-DNA complex structures were generated by docking calculations. The complex structures were reoptimized with the PM6/Mozyme method with implicit solvation in water. The structures were also calculated in the gas phase. The interaction enthalpies between the protein and DNA within the complexes obtained in the PM6/Mozyme with solvation optimized structures were calculated with the single-point PM6-DH2/Mozyme method (PM6 with dispersion, H-bond correction and Mozyme) with solvation. The results supported the specific and non-specific interactions in the complexes obtained from the docking experiments. The binding enthalpies of the specific and non-specific DNA binding to the protein differed significantly. The interactions between the nucleic acid strands in duplexes were also evaluated; these interactions between the base pairs were different because of the different "**G**. . .**C:A**. . .**T**" ratios in the DNA molecules studied. The stacking interactions between the nucleic bases were also characterized in the DNA duplexes.

1. Introduction

Zinc-fingers (ZFs) present the possibility to target specific nucleic acid sequences. These protein domains have been identified in transcription factors, which are highly specific proteins that control gene activation and inactivation in living cells. The general sequence of a ZF domain can be written as follows: (Tyr,Phe)-X-Cys-X2-4-Cys-X3-Phe-X5-Leu-X2-His-X3–5-His [1]. The ZF domain consists of two antiparallel β-sheets connected by a loop to an α-helix. Two cysteines are located in the β-sheet, while the helix contains two histidines, which are necessary to coordinate to the zinc (II) ion (Fig. 1A) in order to form a functional structure. One ZF domain usually recognizes three subsequent bases in a DNA sequence by making specific interactions with the bases within the major groove of the DNA (Fig. 1B). The amino acid residues responsible for specific contacts are at positions −1, 1, 2, 3, 5, and 6, numbered relative to the start of the α-helix [1–3]. By varying these residues in an array of ZF domains, designed ZF proteins are able to recognize a unique sequence even in the DNA of the size of the human genome. ZF proteins can be designed to recognize practically any chosen nucleotide sequence. It was reported that such ZF proteins can also be linked to a catalytically active agent to create specific artificial enzymes [4]. The most relevant examples of the chimeric nucleases are the ZF-FokI fusion proteins, first published by Kim et al. [5]. Since then, a large number of different applications arose from this experiment. Genes in *Drosophila* [6,7], *Xenopus laevis* [8], *Caenorhabditis elegans* [9], Arabidopsis [10], plants [11] and mice [12] were targeted for mutagenesis. Recently, it was found that the specifically positioned chromosomal cleavage by ZF nucleases increased the frequency of homologous recombinations in the human genome [13–16]. Thus, by inducing the cell's own repair mechanism, new or missing genome sequences can be introduced into the cell's genetic material. It gives hope for children born with monogenic diseases, e.g., Duchenne Muscular Dystrophy [17–19], for a cure through gene therapy. The specificity of the DNA binding, i.e., the optimization of the design and selection procedures, is a critical requirement for such prospective applications. Currently, the selection of a new specific ZF array for the recognition of a long DNA sequence is either based on experiments [20,21], the ZF libraries with characterized proteins [2,22–24], or affinity selection methods, such as the phage display [25–27] or the ribosome display [28] methods. A more promising approach is the combination of these methods in a rationally optimized design: first using the database of the Zinc-Finger Tools program package [2], for example, and then using the commonly applied phage display method in a sequential way with a restriction such that only one finger is varied at a time.

* Corresponding author. Tel.: +36 62 544 612; fax: +36 62 312 921.
E-mail address: kortve@chem.u-szeged.hu (T. Körtvélyesi).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.03.002

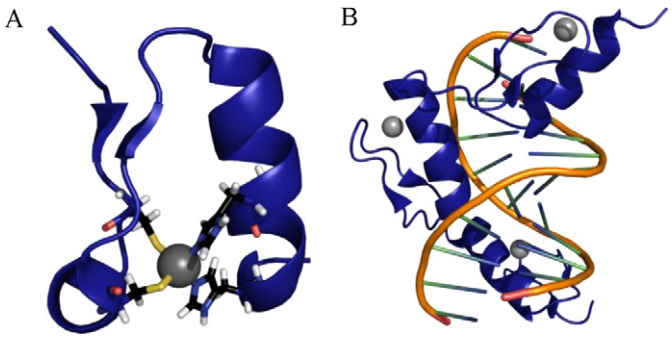

Fig. 1. (A) The structure of a ZF domain from the PDB file 1TF3. The amino acids involved in the metal ion (grey sphere) binding are depicted by sticks. (B) A zinc-finger protein in complex with DNA (1A1K). The protein is represented by blue, the DNA by orange cartoons, while the zinc ions are shown as grey spheres.

To the best of our knowledge, theoretical calculations focusing on the ZF–DNA interactions with large systems have not yet been performed, and no full ZF-structure simulation with quantum chemical methods were found at all. However, several other results are published that focus on the particular properties of the ZF domain. Gresh et al. [29] generated conformers for two zinc-binding sites of the HIV-1 nucleocapsid protein with using a polarizable force field SIBFA (Sum of interactions *Ab Initio* Calculated). The method generates acceptable details on the neighborhood of zinc atoms in ZFs. Dudev et al. [30] performed the all-electron calculations of the nucleation structures in metal-induced zinc-finger folding to find the effect of the backbone on the folding. The redox potential of retroviral ZFs were determined experimentally and with DFT calculations [31]. Carr et al. [32] simulated the ZF structure with a molecular dynamics method. They found that the structure did not deviate significantly from the set of structures determined for the zinc-free peptide.

In the present study, to obtain an impression of the possibility of theoretically modeling the specificity of the interactions, calculations have been performed on specific and nonspecific DNA complexed with a three-ZF-array protein based on semi-empirical approaches. As an ideal target structure to model, we have chosen a ZF-array protein crystallized in the presence of a specifically recognized DNA sequence [33].

2. Calculations

The structure of a ZF protein (1MEY) [33] was obtained from the Protein Data Bank (PDB, www.pdb.org). The original amino acid sequence and the sequence used as the model for the calculation are described in Scheme 1. The structure of the 1MEY model is depicted in Fig. 2. The sequence of the original (specific) [33] and the modified (non-specific) DNA duplexes (**DNA1** and **DNA2**, respectively) are shown in Scheme 2. **DNA2** is a target DNA sequence for a newly designed ZF protein in our experimental work in progress. As it significantly differed from the consensus DNA target of 1MEY protein it was considered to be a non-specific DNA in this study. The H-atoms were added with VEGA ZZ [34] and checked manually. The first calculations were performed to fit the H-atoms.

The non-specific DNA duplex (**DNA2**) was generated as a B-form DNA with X3DNA [35] and aligned with VEGA ZZ [34].

The 1MEY protein was docked to its original and non-specific DNA (Scheme 2) with the program HEX, version 5.1 [36,37] with a spherical polar basis function (SPB) and with a global docking procedure. The complexes were generated by one translation and by 5 rotations around the Euler angles applied in the 3D-FFT fast line method during docking without a bump. In docking, the shape and electrostatic complementarity were considered. The grid dimension was 0.6 Å. The receptor and ligand range was 180°. The step size was 7.5°. The twist range was 360°. The steric and the final scan number was 25 and 25, respectively. The best-docked structures were optimized with the electrostatic and van der Waals interactions of the OPLSA force field built into the package. This method is very sensitive to the initial position of the ligand molecule. If the initial coordinates were far from the target, the best structures had no acceptable values of score function.

1MEY protein sequence [33]:

F1- EKPYKCPECGKSFS**QSSN**L**QK**HQRTHT-
F2-GEKPYKCPECGKSFS**QSSD**L**QK**HQRTHT-
F3-GEKPYKCPECGKSFS**RSDH**L**SR**HQRTHQ

1MEY model sequence:

-MEKPYKCPECGKSFS**QSSN**L**QK**HQRTHT-
-GEKPYKCPECGKSFS**QSSD**L**QK**HQRTHT-
-GEKPYKCPECGKSFS**RSDH**L**SR**HQRTHQ-NKK

Scheme 1. The original (PDB ID: 1MEY), sequence and the model sequence of the 1MEY protein used in the calculations (bold letters show the residues participating in DNA binding).

The best structures were optimized with the PM6 semi-empirical quantum chemical method [38], both with and without Mozyme, as implemented in MOPAC2009 [39]; both in the gas phase and in implicit water solvent. The implicit solvation was performed with COSMO [40], and the dielectric constant ε was 78.4. The gradient norm was accepted at an average of 34 kJ/mol/Å for the complex molecules. The usage of counter ions for DNA and

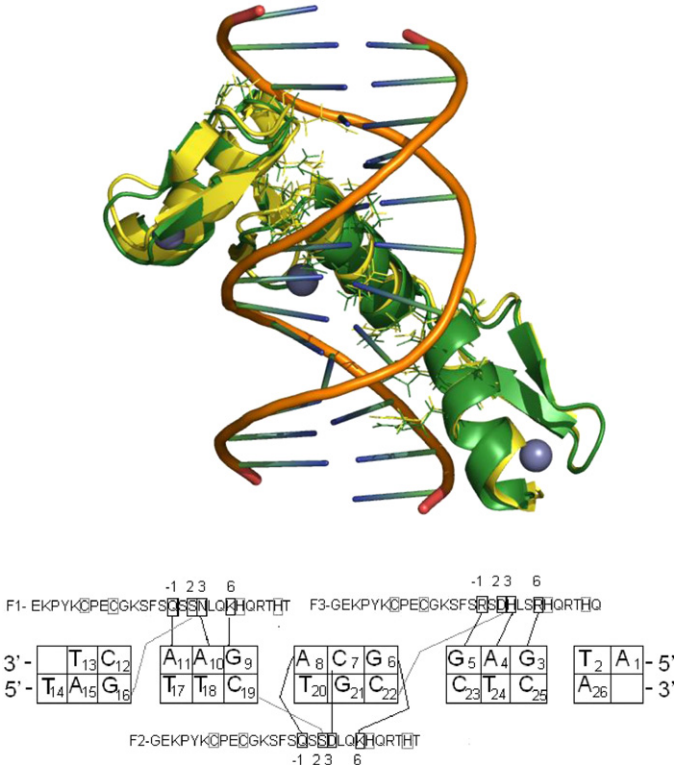

Fig. 2. Interactions between **DNA1** and 1MEY ZF protein.

The sequence of specific **DNA1** to 1MEY protein:

```
      F3 F2 F1
5′-ATGAGGCAGAACT -3′
3′- ACTCCGTCTTGAT-5′
```

The artificially created non-specific DNA (**DNA2**) sequence:

```
(1)           F3 F2 F1
     5′-CCCTGGGGTCACC -3′
     3′- GGACCCCAGTGGC-5′
(2)        F1 F2 F3
```

The consensus recognition sequence of 1MEY protein [33]:

```
        F3      F2      F1
     5′-GGG(G,T)C(A,G)GAA-3′
     3′-CCC(C,A)G(T,C)CTT-5′
```

Scheme 2. The original and the modified DNA (**DNA1** and **DNA2**) sequences. The nucleic bases bound to the ZF are in bold. The consensus DNA sequence from ref. [33] is also described in the scheme. The positions of F1, F2 and F3 finger domains schematically indicate the recognition base triplets in the complexes.

DNA-protein complexes was not possible because of the interaction enthalpy calculations in the complexes. The structures were optimized with the LBFGS method. The binding enthalpies were calculated for the docked and optimized structures with PM6-DH2/Mozyme/W//PM6/Mozyme/W [41,42], meaning that the energetic and electronic structures of the molecules were optimized with PM6/Mozyme in water (W), and the final structure was applied for the calculation of the energetic and electronic structures with PM6-DH2/Mozyme in water. No frequency calculations were carried out because of the following reasons: (i) the initial geometries were partly built up on the basis of XRD structure with ca. 700 atoms, (ii) with Mozyme and (iii) COSMO methods no frequency calculations were suggested [43].

The enthalpies for the protein–DNA interactions have been calculated as follows:

$$\Delta H^{\circ}(\text{interaction}) = \Delta_f H^{\circ}(\text{complex}) - \Delta_f H^{\circ}(\text{protein}) - \Delta_f H^{\circ}(\text{duplex}) \quad (1)$$

The optimized DNA structures were also modified by deleting the backbone atoms and adding hydrogen atoms to the remaining –CH– to make CH_3, and the H-atoms were optimized again. From this optimization, we calculated the heat of formation for the duplexes of **mG**. . .**mC** and **mA**. . .**mT** (where m is the methyl group). The interaction enthalpies ($\Delta_f H^{\circ}$) were predicted by the following expression:

$$\Delta H^{\circ}(H\text{-bonds}) = \Delta_f H^{\circ}(\text{duplex}) - \Delta_f H^{\circ}(\text{S1}) - \Delta_f H^{\circ}(\text{S2}) \quad (2)$$

where S1 and S2 are the single strand 1 and the single strand 2 of the duplexes, respectively. The calculation of the sum of π–π stacking, backbone deformation and backbone-nucleic acid interactions were calculated by Eq. (3):

$$\Delta H^{\circ}(\pi\text{–}\pi \text{ stacking}) = \Delta H^{\circ}(\text{H-bonds, duplex}) - i\Delta H^{\circ}(\text{H-bonds, mG} \ldots \text{mC}) - j\Delta H^{\circ}(\text{H-bonds, mA} \ldots \text{mT}) \quad (3)$$

Table 1
Interactions of **DNA1** with three ZF domains (F1,F2 and F3) of 1MEY protein [33].

1MEY + Zn + DNA1			
ZF domain	Position of AA	AA	Interacting base
F1	−1	Q16	A11
F1	1	S17	–
F1	2	S18	G16
F1	3	N19	A10
F1	5	Q21	–
F1	6	K22	G9
F2	−1	Q44	A8
F2	1	S45	–
F2	2	S46	C19
F2	3	D47	C7
F2	5	Q49	–
F2	6	K50	G6
F3	−1	R72	G5
F3	1	S73	–
F3	2	D74	C22
F3	3	H75	A4
F3	5	S77	–
F3	6	R78	G3

where *i* and *j* are the number of the **mG**. . .**mC** and **mA**. . .**mT** base pairs. The heat of formation of the **mG**. . .**mC** and **mA**. . .**mT** base pairs were calculated (i) in ideally optimized structures and (ii) separated from the duplexes without reoptimization. The stacking interactions were also calculated with another method:

$$\Delta H^{\circ}(\pi - \pi \text{ stacking}) = \Delta_f H^{\circ}(\text{m-duplex}) - i\,\Delta_f H^{\circ}(\text{mG} \ldots \text{mC}) - j\Delta_f H^{\circ}(\text{mA} \ldots \text{mT}) - \sum \Delta_f H^{\circ}(\text{mB without pair}) \quad (4)$$

where *i* and *j* are the number of the **mG**. . .**mC** and **mA**. . .**mT** base pairs, as before, and mB is the methylated base without a pair, m-duplex is the duplex without backbone, methylated on the bases.

3. Results and discussion

The 1MEY [33] model from the XRD crystal structure was the initial target in our docking calculations (for the sequence of the model see Scheme 1). The design of the 1MEY protein was based on the Sp1 consensus sequence [44], changing only the amino acid residues that interacted with specific bases in the DNA, while keeping the rest constant in the individual ZF domains. This approach allows for the design of new proteins with the same framework, recognizing arbitrary DNA sequences. Therefore, it is an ideal target for modeling studies intended to aid the design of ZF specificity.

The crystal structure of the 1MEY protein with its cognate DNA revealed the interactions between the amino acids in positions −1, 2, 3, and 6, numbered relative to the start of the α-helix in detail. The most important conclusions concerning the binding sites were the following (see also Table 1 and Fig. 2):

(i) In position −1, the glutamine can interact with more than one base because of its dual H-donor and acceptor property. It binds adenine and guanine or cytosine at the 3′ end of the binding site. The arginine in the same position (third ZF domain of 1MEY protein) binds the guanine base stabilized by the aspartic acid in position 2 by a salt bridge.
(ii) In position 2, serines (first and second ZF domains) do not take part in site-specific interactions similar to the aspartic acid in the third ZF domain. Rather, they stabilize the complex upon interacting with the complementary DNA strand at the 3′ end outside the recognition base triplet.
(iii) In position 3, all three ZF domains contain different amino acids, but all of them interact through hydrogen bonds with

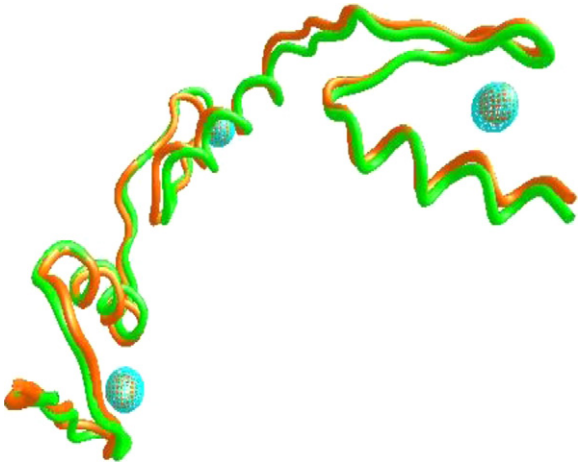

Fig. 3. 1MEY ZF protein backbone structure (without DNA) as obtained by XRD [33] (green) and the PM6/Mozyme/W optimized structure (red). (For interpretation of the references to color in this sentence, the reader is referred to the web version of the article.)

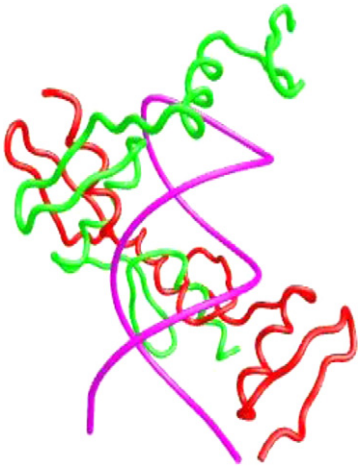

Fig. 4. The backbone structures of the best 1MEY protein **DNA2** complexes obtained in the docking experiments: (1) green and (2) red. (For the further explanation see text.) (For interpretation of the references to color in this sentence, the reader is referred to the web version of the article.)

the middle bases in their recognition base triplets. These interactions are the dominating specificity determinants.

(iv) The interactions found for the amino acid side-chains in position 6 are largely affected by the previous binding scheme. The lysines bind strongly to the N and/or O donor atoms of a guanine, while the arginines form a dual H-bond with a guanine base.

The model of the 1MEY protein was obtained and optimized as described above. The difference between the XRD and the PM6-optimized structure (backbone without the side chains) is shown in Fig. 3. The RMSD was found to be 0.38 Å.

The results of HEX [36] calculations with the total energies considering the potential for the shape, electrostatic and van der Waals potentials, as well as the number of bumps are summarized in Table 2. In the initial structures, the distance between the center of mass of the DNA and the protein was varied. We found that the larger the distance, the worse the results. A distance of 0.5–1.0 Å was chosen and accepted for the calculations. **DNA1** fitted to the 1MEY protein similarly, as it was experimentally shown (see Fig. 2) with an acceptable value of the score function. **DNA2** is an artificially created non-specific ligand for the 1MEY protein. Because the DNA is mainly in its B form in the complexes with ZF proteins with a small distortion of the major groove accommodating the ZF protein chain [3,33], we also applied this form for **DNA2**.

Two significant structures have been observed during the docking experiments:

(1) In the complex with the best score function, **DNA2** shifted compared to the **DNA1** position in the XRD structure (see Fig. 4). Upon this movement, the **DNA2** sequence, including the region between G07-C12 and A19-C14, has been fitted to the 1MEY protein. The analysis of a number of crystal structures of cys2-his2-type zinc-finger (ZF) proteins complexed with their specific DNA suggested the modular behavior of the ZF arrays [3]. This is well represented by the rearrangement of the 1MEY protein–**DNA2** complex upon docking because a consensus binding site [33] (see Scheme 2) was found for fingers 2 and 3 within the **DNA2** sequence. Finger 1 was loosely bound to the end of the DNA strands. This interaction is a partially specific one.

(2) In the second structure, the 1MEY protein remained in the major groove of **DNA2**, interacting with T04, G06-T09, A11 and A24-T17 bases (see Scheme 3). This is a similar structure to the XRD structure for the 1MEY protein with **DNA1** but presumably with weaker interactions than in the specific complex. This site is characteristic for non-specific binding of **DNA2**.

In the next step, the complex structures obtained in the docking procedure were calculated with PM6-DH2/Mozyme/W//PM6/

Table 2
The results of **DNA1** and **DNA2** docking to 1MEY by HEX[a].

Model name	E_{total}	E_{shape}	E_{force}	V_{clash}	Bmp
1MEY + Zn + DNA2 (1)	−566.4	−455.0	−111.4	1.0	0
1MEY + Zn + DNA2 (2)	−425.7	−473.6	47.8	11.8	0
1MEY + Zn + DNA1	−599.5	−571.6	−27.9	0.6	0
1MEY + DNA2 (1)	−559.8	−464.3	−95.5	4.1	0
1MEY + DNA2 (2)	−440.8	−467.0	26.3	11.2	0
1MEY + DNA1 dist 6 Å	309.8	−201.1	511.0	37.0	0
1MEY + DNA1 dist 4 Å	−7.0	−177.2	170.2	15.1	0
1MEY + DNA1 dist 3 Å	−293.4	−462.3	168.9	6.8	0
1MEY + DNA1 dist 2 Å	−346.9	−481.9	135.0	0.0	0
1MEY + DNA1 dist 1 Å	−589.6	−571.1	−18.5	0.0	0
1MEY + DNA1 dist 0.5 Å	−589.6	−571.1	−18.5	0.0	0
1MEY + DNA1	−597.3	−568.4	−28.9	1.1	0
1MEY + DNA1 dist −1 Å	−547.7	−540.6	−7.0	10.9	0

[a] E_{total}: total score energy; E_{shape}: energy charaterizing the shape complementarity; E_{force}: interaction energy; V_{clash}: value of the penalty function of the clashes; Bmp: number of bumps.

(1) Non-specific binding interactions of **DNA2** with 1MEY (the interaction is specific for the first two ZF where it shifted):

```
5′-C01 C02 C03 T04 G05 G06 G07 G08 T09 C10 A11 C12 C13 -3′
3′-    G26 G25 A24 C23 C22 C21 C20 A19 G18 T17 G16 G15 C14-5′
```

Interactions:

```
F1 K22-G15; N19-G16,C14,C12; S18-C14, Q16-C14
F2 K50-G18; D47-C10, S46-T17; Q44-A11,C10
F3 R78-G7; H75-G7; D74-G18,A19; S73-G18; R72-A19,G8
```

(2) Non-specific binding of **DNA2** interactions with 1MEY:

```
5′-C01 C02 C03 T04 G05 G06 G07 G08 T09 C10 A11 C12 C13 -3′
3′-    G26 G25 A24 C23 C22 C21 C20 A19 G18 T17 G16 G15 C14-5′
```

Interactions:

```
F1 Q16-T4,A24; S18-T4; N19-C23; K22-C22,G6
F2 Q44-G7,C21; S46-G6; D47-C20; K50-G8,A19
F3 R72-G18,T9; S73-G8,A19; D74-T9,G8; H15-T17; R78-G18
```

Scheme 3. The interactions within the two significant structures (1) and (2), obtained from the docking experiments with the non-specific **DNA2** to 1MEY protein. The nucleic bases interacting with the ZF protein are shown in bold.

Table 3
Interaction enthalpies (kJ/mol) between 1MEY, **DNA1** and DNA2 (type (1)) calculated by PM6-DH2/Mozyme/W//PM6/Mozyme/W [45].

	1MEY-DNA1 interaction	**DNA1**(S1)-**DNA1**(S2) interaction	1MEY-DNA2 type (1) interaction	**DNA2**(S1) + **DNA2**(S2)interaction
Without Zn	−1905	−440	−1172	−594
With Zn	−1627	−456	−1217	−572

Mozyme/W (the gas phase calculations gave unreliable interaction data because of the repulsion effect of the charged groups). The optimized complexes were separated, and single-point calculations were performed to obtain the formation enthalpies of the protein and the DNA strands, respectively.

The results calculated on the basis of Eq. (1) are summarized in Table 3 with and without Zn(II) (in the calculations performed without Zn(II), cys with CH_2-SH was used). The interaction enthalpies between 1MEY and **DNA1** or **DNA2** calculated for the best structures obtained from the docking experiments supported that the 1MEY-ZF protein bound preferably to its original DNA sequence (**DNA1**) (by −1905 kJ/mol without Zn(II) and −1627 kJ/mol with Zn(II) in the 1MEY-ZF protein). At the same time, with non-specific **DNA2**, the type (1) binding mode (see above) yielded a stronger interaction than for the type (2) complex, while the calculated interaction enthalpy was smaller than for the specific complex (see Table 3). The differences between the specific and non-specific complexes may be considered significant. This fact can be explained by the extensive amino acid side-chain-DNA strand contacts (including the interactions with the DNA backbone) formed by the ZF protein bound to the major groove of the DNA (see Scheme 3).

No significant difference was found in the interaction enthalpies of the DNA duplexes (**DNA1S1**. . .**DNA1S2** and **DNA2S1**. . .**DNA2S2**) in the calculation with and without Zn(II) (−440 to −456 kJ/mol and −594 to −572 kJ/mol, respectively). This indicates that the metal ion affects the protein but not the DNA.

It is also worth noting the results concerning the interactions between the DNA single strands. By deleting the backbone in **DNA1** and **DNA2**, replacing –CH– with CH_3 and optimizing the H-atoms in the molecule, the interaction enthalpies were obtained from single-point calculations. The Watson-Crick (WC) interaction enthalpies of the base pairs **mA**. . .**mT** and **mG**. . .**mC**, calculated with different PM6 methods, are summarized in Table 4. The interaction enthalpies obtained from our calculations in gas phase were

Table 4
Interaction enthalpies (kJ/mol) between **mA**. . .**mT** and **mC**. . .**mG** methyl-nucleic bases by different theoretical methods.

Methods	**mA**. . .**mT**	**mC**. . .**mG**
PM6	−37.1	−76.6
PM6/W[a]	−1.1	−3.7
PM6/M	−37.1	−76.6
PM6/M/W	−1.2	−3.1
PM6/DH2//PM6	−66.7	−120.7
PM6/M/DH2//PM6/M[b]	−60.8	−120.9
PM6/W/DH2//PM6/W	−29.6	−52.5
PM6/M/W/DH2//PM6/M/W	−29.6	−52.8
PM6/DH2	−69.7	−117.1
PM6/W/DH2	−34.7	−58.9
PM6/M/DH2	−69.7	−117.2
PM6/M/W/DH2	−34.5	−52.0

[a] W: dielectric constant is 78.4.
[b] M: Mozyme, DH2: DH2 [45].

similar to the different level of *ab initio* and density functional (DFT) methods. Jurecka and Hobza [45] published −108.02 kJ/mol and −59.87 kJ/mol (RI-MP2/TZVPP) for the WC H-bond in the gas phase for the interaction between **mG**. . .**mC** and **mA**. . .**mT** base pairs, respectively. The WC interaction energies are −100.7 and −46.8 kJ/mol for **mG**. . .**mC** and **mA**. . .**mT**, respectively, calculated by DFTB+ with H-bond and dispersion corrections [46] in the gas phase. The interaction enthalpies of **mG**. . .**mC** and **mA**. . .**mT** are −117.1 kJ/mol and −69.7 kJ/mol in the gas phase and −58.9 kJ/mol and −34.7 kJ/mol in water, respectively, calculated with PM6-DH2. The interaction enthalpies in water are significantly smaller.

The interaction enthalpies were also calculated with PM6-DH2/Mozyme/W//PM6/Mozyme/W, which were −52.8 kJ/mol and −29.6 kJ/mol for **mG**. . .**mC** and **mA**. . .**mT**, respectively. Considering the heats of enthalpy of the two strands, the interaction enthalpies for **DNA1** and **DNA2** were calculated to be −596.5 kJ/mol and −738.9 kJ/mol, respectively. On average, the interaction enthalpies between the base pairs were −49.7 kJ/mol and −61.6 kJ/mol in **DNA1** and **DNA2**, respectively. The interaction enthalpies thus far were determined in single base pairs or in small structures with two or three base pairs in the gas phase. Our experience was that the electronic structures and the charges were completely different among bases; base pairs with H-bonds and in systems including more base pairs. This clearly affected the strength of the H-bonds and π–π stacking. With the values of the interaction enthalpies in **mG**. . .**mC** and **mA**. . .**mT** (see Eq. (2)), the enthalpies for π–π stacking were calculated with PM6-DH2/Mozyme/W//PM6/Mozyme/W for **DNA1** (5A, 2T, 4G, 2C. . .2A, 5T, 4C, 2G) and **DNA2** (6C, 1A, 2T, 4G. . .5G, 2A, 5C, 1T), and we obtained −102.4 kJ/mol and −175.3 kJ/mol in the duplex including twelve base pairs (Eq. (3)), respectively. (The duplexes exert unpaired bases at their 5′ ends – see Scheme 2). The stacking enthalpies per base pair were found on average in **DNA1** and **DNA2** to be −8.5 kJ/mol and −14.6 kJ/mol, respectively. The difference is due to the difference in the number of **mG**. . .**mC** and **mA**. . .**mT** base pairs. The stacking enthalpies appeared to be too small. The other values for stacking enthalpies were calculated by Eq. (4) (with PM6-DH2/Mozyme/W//PM6/Mozyme/W for **DNA1** (5A, 2T, 4G, 2C. . .2A, 5T, 4C, 2G) and **DNA2** (6C, 1A, 2T, 4G. . .5G, 2A, 5C, 1T)), and the results of −337.6 kJ/mol (−459.4 kJ/mol considering the heat of formation of the real base pair structure) and −281.8 kJ/mol (−436.0 kJ/mol considering the heat of formation of the real base pair structure), respectively, were obtained, which were on average −28.1 kJ/mol (−41.8 kJ/mol considering the heat of formation of the real base pair structure) and −23.5 kJ/mol (−39.6 kJ/mol considering the heat of formation of the real base pair structure) per base pair, respectively.

After considering only the stacking in one strand, we also calculated the stacking enthalpies. In DNA1 S1 (5A, 2T, 4G, 2C) and S2 (2A, 5T, 4C, 2G), the stacking enthalpies were −105.5 kJ/mol (−8.8 kJ/mol for one mB) and −130.5 kJ/mol (−10.9 kJ/mol for one mB), respectively. The same values in DNA2 S1 (6C, 1A, 2T, 4G) and S2 (5G, 2A, 5C, 1T) were −17.4 kJ/mol (−1.5 kJ/mol for one mB) and −90.1 kJ/mol (−7.5 kJ/mol for one mB), respectively. The sum of the stacking enthalpies of the DNA strands is different from that obtained by Eq. (4).

Based on previous publications, the above results and the (Mulliken) charge calculations for the bases alone, base pairs, and for two-three or more base pairs, we can conclude that the charges were different in different systems [47]. This means that the values calculated by Eq. (4) can be accepted as the more realistic result. No acceptable interaction enthalpy characteristics for the stacking interactions were obtained in the single DNA strands, supporting the reorganization of the electron distribution after obtaining base pairs with H-bonds [47].

4. Conclusions

Semi-empirical quantum chemical calculations were performed to model specific and non-specific binding interactions between a 1MEY-ZF protein and DNA. As a result of docking experiments, we found the specific binding site for the original DNA (**DNA1**) as in the XRD structure. The calculations with a designed non-specific DNA (**DNA2**) revealed two different structures for the 1MEY protein-DNA complex: (1) a specific binding site for fingers 2 and 3 have been found in the sequence of **DNA2** and (2) a non-specific binding site was found in the structure similar to the XRD structure of the 1MEY protein with **DNA1**. The calculated interaction enthalpies correlated well with the specificity of the ZF protein–DNA interactions, but further calculations must be performed to validate a reliable scale of interaction enthalpies *vs*. specificity of the interaction. For the first time in the literature, the calculations provided values for the H-bond and π–π stacking interactions for large DNA duplexes, as determined from quantum chemical calculations considering the polarizability. We found that the change in the charges, depending on the geometry, must be considered in the simulation of the DNA duplexes.

Acknowledgements

This work has received support through the Hungarian Science Foundation (OTKA-NKTH CK80850, OTKA K72781 and K61577) as well as through TAMOP-4.2.1./B-09/1. The authors are grateful for the use of computational resources at the HPC group of University Szeged.

References

[1] D. Jantz, B.T. Amann, G.J. Gatto Jr., J.M. Berg, Chem. Rev. 104 (2004) 789–799.
[2] J.G. Mandell, C.F.III Barbas, Nucleic Acids Res. 34 (2006) W516–W523.
[3] N.P. Pavletich, C.O. Pabo, Science 252 (1991) 809–817.
[4] T. Nakatsukasa, Y. Shiraishi, S. Negi, M. Imanishi, S. Futaki, Y. Sugiura, Biochem. Biophys. Res. Commun. 330 (2005) 247–252.
[5] Y.-G. Kim, J. Cha, S. Chandrasegaran, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 1156–1160.
[6] K. Beumer, G. Bhattacharya, M. Bibikova, J.K. Trautman, D. Carroll, Genetics 172 (2006) 2391–2403.
[7] M. Bibikova, M. Golic, K.G. Golic, D. Carroll, Genetics 161 (2002) 1169–1175.
[8] M. Bibikova, D. Carroll, D.J. Segall, J.K. Trautman, J. Smith, Y.-G. Kim, S. Chandrasegaran, Mol. Cell. Biol. 21 (2001) 289–297.
[9] J. Morton, M.W. Davis, E.M. Jorgensen, D. Carroll, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 16370–16375.
[10] M.A. Lloyd, C.L. Plaisier, D. Carroll, G.N. Drews, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 2232–2237.
[11] D.A. Wright, J.A. Townsend, R.J. Winfrey Jr., P.A. Irwin, J. Rajagopal, P.M. Lonosky, B.D. Hall, M.D. Jondle, D.F. Voytas, Plant J. 44 (2005) 693–705.
[12] M. Mani, K. Kandavelou, F.J. Dy, S. Durai, S. Chandrasegaran, Biochem. Biophys. Res. Commun. 335 (2005) 447–457.
[13] A. Pingoud, G.H. Silva, Nat. Biotechnol. 25 (2007) 743–744.
[14] A. Lombardo, P. Genovese, C.M. Beausejour, S. Colleoni, Y.L. Lee, K.A. Kim, D. Ando, F.D. Urnov, C. Galli, P.D. Gregory, M.C. Holmes, L. Naldini, Nat. Biotechnol. 25 (2007) 1298–1306.
[15] F.D. Urnov, J.C. Miller, Y.-L. Lee, C.M. Beausejour, J.M. Rock, S. Augustus, A.C. Jamieson, M.H. Porteus, P.D. Gregory, M.C. Holmes, Nature 435 (2005) 646–651.
[16] E.A. Moehle, J.M. Rock, Y.-L. Lee, Y. Jouvenot, R.C. DeKelver, P.D. Gregory, F.D. Urnov, M.C. Holmes, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 3055–3060.
[17] G. Cossu, M. Sampaolesi, TRENDS Mol. Med. 10 (2004) 516–520.
[18] J.S. Chamberlain, Nature 444 (2006) 552–553.
[19] K.J. Nowak, K.E. Davies, EMBO Rep. 5 (2004) 872–876.
[20] W. Nomura, Y. Sugiura, Methods Mol. Biol. 352 (2007) 83–93.
[21] M. Dhanasekaran, Sh. Negi, Y. Sugiura, Acc. Chem. Res. 39 (2006) 45–52.
[22] D.J. Segall, C.F. Barbas III, Curr. Opin. Biotechnol. 12 (2001) 632–637.
[23] D. Carroll, J.J. Morton, K.J. Beumer, D.J. Segal, Nat. Protocols 1 (2006) 1329–1341.
[24] M.L. Maeder, S. Thibodeau-Beganny, A. Osiak, D.A. Wright, R.M. Anthony, M. Eic htinger, T. Jiang, J.E. Foley, R.J. Winfrey, J.A. Townsend, E. Unger-Wallace, J.D. Sander, F. Müller-Lerch, F. Fu, J. Pearlberg, C. Göbel1, J.P. Dassie, S.M. Pruett-Miller, M.H. Porteus, D.C. Sgroi, A.J. Iafrate, D. Dobbs, P.B. McCray Jr., T. Cathomen, D.F. Voytas, J.K. Joung, Mol. Cell 31 (2008) 294–301.
[25] H.A. Greisman, C.O. Pabo, Science J.C.A.M.D. 275 (1997) 657–661.
[26] A.C. Jamieson, H. Wang, S.-H. Kim, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 12389–12834.
[27] E.J. Rebar, H.A. Greisman, C.O. Pabo, Methods Enzymol. 267 (1996) 129–149.

[28] H. Ihara, M. Mie, H. Funabashi, F. Takahashi, T. Sawasaki, Y. Endo, E. Kobatake, Biochem. Biophys. Res. Commun. 345 (2006) 1149–1154.
[29] N. Gresh, P. Derreumaux, J. Phys. Chem. B 107 (2003) 4862–4870.
[30] T. Dudev, C. Lim, J. Am. Chem. Soc. 129 (2007) 12497–12504.
[31] I.A. Topol, C. McGrath, E. Chertova, C. Dasenbrock, W.R. Lacourse, M.A. Eissenstat, S.K. Burt, L.E. Henderson, J.R. Casas-Finet, Prot. Sci. 10 (2008) 1434–1445.
[32] M.D. Carr, A. Pastore, H. Gausepohl, R. Franck, P. Roesch, J. Eur. Biochem. 188 (1990) 455–461.
[33] C.A. Kim, J.M. Berg, Nat. Struct. Biol. (1996) 940–945.
[34] A. Pedretti, L. Villa, G. Vistoli, 18 (2004) 167–173.
[35] W.K. Olson, M. Bansal, S.K. Burley, R.E. Dickerson, M. Gerstein, S.C. Harvey, U. Heinemann, X.-J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C.-S. Tung, E. Westhof, C. Wolberger, H.M. Berman, J. Mol. Biol. 313 (2001) 229–237.
[36] http://www.loria.fr/~ritchied/hex.
[37] D.W. Ritchie, D. Kozakov, S. Vajda, Bioinformatics 24 (2008) 1865–1873.
[38] J.J.P. Stewart, J. Mol. Mod. 13 (2007) 1173–1213.
[39] J.J.P. Stewart, Mopac2009, http://www.openmopac.net/MOPAC2009.html.
[40] A. Klamt, G. Schüümann, J. Chem. Soc. Perkin Trans. 2 (1993) 799–805.
[41] M. Korth, M. Pitonák, J. Rezác, P. Hobza, J. Chem. Theory Comp. 6 (2010) 344–352.
[42] J. Rezác, J. Fanfrlik, D. Salahub, P. Hobza, J. Chem. Theory Comp. 5 (2009) 1749–1760.
[43] MOPAC2009 Manual, www.openmopac.net (keywords EPS and Mozyme).
[44] D.J. Segal, R.R. Beerli, P. Blancafort, B. Dreier, K. Effertz, A. Huber, B. Koksch, C.V. Lund, L. Magnenat, D. Valente, Biochemistry 42 (2003) 2137–2148.
[45] P. Jurecka, P. Hobza, J. Am. Chem. Soc. 125 (2003) 15608–15613.
[46] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert, Phys. Rev. B 58 (1998) 7260–7268.
[47] T. Körtvélyesi, E.A. Hoffmann, I. Pálinkó, G. Nagy, B. Gyurcsik, B. Penke, prepared for publication.