**STATISTICAL RELATION BETWEEN THE POSITIONS OF THE ALPHA-HELIX IN THE ZINC-FINGER DNA-BINDING DOMAIN: RESULTS FROM THE PHAGE DISPLAY DATA ANALYSIS.**

AFONNIKOV D.+, WINGENDER E.1

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, 10 Lavrentiev Ave., Novosibirsk, 630090, Russia;
e-mail: ada@bionet.nsc.ru
1Molecular Bioinformatics of Gene Regulation, GBF, Braunschweig, Germany
e-mail: ewi@gbf.de

+Corresponding author

Keywords: transcription factor, zinc finger, amino acid substitutions, correlations, DNA-binding

**1. Introduction**

The C2H2 zinc finger domain is one of the major classes of the DNA-binding motifs [1, 2]. Its tertiary structure consists of two beta-sheets packed against alpha-helix. The alpha-helix binds into the DNA major groove and includes residues (positions -1, 2, 3, and 6 relative to alpha-helix first residue), which are in specific contact with the nucleotide bases [3]. However, the details of this specific recognition remain unclear.

In this report we present the results of correlation analysis of the previously published data on the Zif268 polypeptide phage display experiments [4, 5]. These data represent the set of amino acid sequences of the alpha-helical region, which were selected on the basis of the best DNA-binding affinity from a large pool of possible polypeptide mutants. Note that these "artificially selected" polypeptides has no "evolutionary relations" unlike the sequence data collected from the protein databases.

**2. Materials and methods**

***2.1. Sequence data***

The sample included 45 aligned sequences taken from the work of Choo & Klug [4] with pairwise sequence difference at least in one position out of 10. Conservative positions with only one or two different amino acid residue types in the sequence alignment has been excluded from the analysis. As a result, we have considered pairwise correlations between 7 positions of the alpha-helix (-1, 1, 2, 3, 5, 6, and 8).

***2.2. Estimation of the pairwise positional dependence***

To estimate the mutual dependence between the amino-acid substitutions at the positions of the zinc-finger DNA-binding helix, we used the approach reported earlier [6]. This approach estimates the partial correlation coefficient between the values of a certain physico-chemical characteristic observed at the positions of proteins from the sample analyzed. This allowed us to estimate the correlation between the pair of positions that eliminates the influence of the residues at the rest protein positions. In this work we analyze four amino acid characteristics: isoelectric point value [7], hydrophobicity [8], polarity [9], and side-chain volume [10].

***2.3. Estimation of the critical value of the partial correlation coefficients***

The significance threshold for the absolute value of pair partial correlation coefficients has been estimated as described in [6]. For the sample of 45 sequences with sequence length of 7 aa, it was estimated as 0.40 (at a 99% significance level).

**3. Results and discussion**

Our analysis discovered several significantly correlated position pairs of the zinc-finger DNA-binding helix. These significant correlations have been revealed for the three physico-chemical quantities (isoelectric point, side chain volume, and hydrophobicity). No significant correlations (at a level of 99%) have been observed for the polarity values. These results allows us to suppose that the most important interactions for the residues of the DNA-binding helix are electrostatic (characterized by an isoelectric point quantity), steric (side chain volume), and hydrophobic. The strongest correlations has been observed for the isoelectric point values (5 significantly correlated pairs out of 21 possible). There are two significantly correlated pairs for the hydrophobicity values and one pair for the side chain volume. Note that all the correlated pairs mentioned above have negative correlation coefficients indicating the compensatory character of the amino acid substitutions at these positions (in terms of the analyzed physico-chemical quantities). The results of the calculations for the isoelectric point and side chain volume quantities are shown in Fig.1.



Figure 1. The diagram of the partial correlation coefficient matrices calculated for the isoelectric point (above the main diagonal) and the side chain volume (below the main diagonal) quantities. The correlation coefficient for each pair is shown as a pattern shaded according to its value (see right upper corner of the diagram for the pattern designation).

For the side chain volume, the negative dependence is observed for the pair of positions -1 and 6. The negative value of correlation coefficients means that if the substitution at the position -1 leads to the decrease in the residue size at this position, the most probable substitution at the position 6 will be for the residue with the larger size. It is interesting that this dependence is in agreement with the hypothesis of Suzuki et al. on the principles of the zinc-finger--DNA recognition [11]. According to this hypothesis, the observed correlation is likely to reflect the effect of orientation of the binding helix in the DNA major groove.

The correlation diagram also shows that the isoelectric point values at the positions -1, 1, 2, 3, and 6 depend negatively on each other. The residues at these positions (except position 1) are known to be DNA base-contacting. Interestingly, the existence of the correlated amino acid substitutions at the position pairs -1, 2 and -1, 3 has been shown recently [12]. It has been suggested that these correlations can be responsible for the DNA-binding specificity of the zinc-finger domain.

The correlations of the isoelectric point values at the positions -1, 1, 2, 3, and 6 show clearly the compensating character of the residue substitutions at these positions. This fact allows us to suppose that the sum of the isoelectric point quantities at these position (Qs) can be approximately invariant for the zinc-finger helices in the sample analyzed. To check this hypothesis, we calculated the Qsvalue for each sequence in our sample and used the sample dispersion of the Qs value (D(Qs)) as a measure of its conservation. In case the positions were independent, the dispersions of sum of the isoelectric point quantities Qs at these positions would expected to be equal to the sum of the positional dispersions of this quantity. The value of D(Qs) in the original sample was estimated as 6.5. This value is significantly lower than the sum of positional dispersions of this quantity (24.91). We also perform a Monte Carlo simulations to estimate the degree of Qsconservation. We generate 100 000 samples by random permutation of amino acids in the columns of the original sequence alignment. This procedure allows the generation of a number of samples with completely independent amino acid substitutions at sequence positions. Then we compare the distribution of the Qsdispersion values Drand(Qs) in these random samples with the value of D(Qs) in original sample. The results of calculations are shown in Figure 2.



Figure 2. The distribution of the Qsdispersion values in the set of random samples (mean value 24.4, standard deviation 4.5). The value of D(Qs) in original sample is indicated by arrow.

It is evident from the diagram that the value of D(Qs) for the original sample is about four-fold standard deviation of the mean value for the simulated samples with independent substitutions. This value is extremely low, and no random sample is found (out of 105 generated) with the value of Drand(Qs) equal to or lower than the value of D(Qs) in the original sample.

These results evidenced that the sum of the isoelectric point quantities at the positions -1, 1, 2, 3, and 6 of the recognition helix in the zinc fingers selected from the large pool of possible mutants on the basis of high DNA-binding affinity is almost invariant.

The sum of the isoelectric point values at the positions mentioned can characterize the total charge at these positions. Therefore, we suppose that the results obtained reflect the conservation of the total charge at the DNA-binding surface of the zinc-finger recognition helix. This electrostatic characteristic is likely to be important for the zinc-finger-DNA interactions and subjected to certain restrictions during phage display selection.

The results of our analysis confirm the existence of pair correlations in amino acid residue substitutions at the positions of the zinc-finger domain. We believe that this correlations reflect the interactions between zinc-finger residues and between zinc-finger and DNA. Therefore, these results can serve as an additional source of information to understand the mechanism of zinc-finger-DNA binding.

**Acknowledgements**

This work was supported by NATO (grant No. HTECH.LG 951149) and German Ministry of Education, Science, Research, and Technology (grant No. X224.6), RFBR (96-04-50006, 97-04-49740) and Russian National Human Genome Project (12312 GCh-5 grants).

**References**

1. J. Miller, A.D. McLachlan and A. Klug, "Repetitive zinc-binding domains in the protein transcription factor IIIA from*Xenopus*oocytes", EMBO J.,**4**, 1609-1614 (1985).
2. E. Wingender, "Classification scheme of eucariotic transcription factors", Mol. Biol. Transl.,**31**, 483-498 (1987).
3. N.P. Pavletich and C.O. Pabo, "Zinc finger - DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A ", Science,**252**, 809-817 (1991).
4. Y. Choo and A. Klug, "Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage", Proc. Natl. Acad. Sci. USA,**91**, 11163-11167 (1994).
5. Y. Choo and A. Klug, "Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions", Proc. Natl. Acad. Sci. USA,**91**, 11168-11172 (1994).
6. D.A. Afonnikov, Yu.V. Kondrakhin, I.I. Titov, "Revealing of correlated positions of the DNA-binding region of the CREB and AP-1 transcription factor families", Molec. Biol.(Russian),**31**, 731. (1997)
7. White, P. Handler, E.L. Smith, R.L. Hill, I.R. Lehman, "Principles of Biochemistry, vol.1.", McGraw-Hill, Inc. (1978)
8. D. Eisenberg, R.M. Weiss, T.C. Terwilliger, "The hydrophobic moment detects periodicity in protein hydrophobicity". Proc. Natl. Acad. Sci. USA,**81**, 140-144 (1984).
9. P.K. Ponnuswamy, M. Prabhakaran, P. Manavalan, " ", Biochim. Biophys. Acta 1980. V. 623. P. 301-316
10. Chothia, "Principles that determine the structure of proteins", Ann. Rev. Biochem.,**53**, 537-572 (1984).
11. M. Suzuki, M. Gerstein and N. Yagi, "Stereochemical basis of the DNA recognition by Zn fingers", Nucl Acid Res,**22**, 3397-3405 (1994).
12. J.R. Desjarlais and J.M. Berg, "Redesigning the DNA-binding specificity of a zinc finger protein: a data base-guided approach", Proteins: Struct. Funct. Genet.,**12**, 101-104 (1992); Ibid.,**13**, 272 (1992).