Neuroendocrinology Letters No.3 June Vol.26, 2005
 ISSN 0172–780X **www.nel.edu**

ORIGINAL ARTICLE

DNA double-helix binds regulatory peptides similarly to transcription factors

Vladimir Khavinson[1], Larisa Shataeva[2] & Anna Chernova[3]

[1] St.Petersburg Institute of Bioregulation and Gerontology, North-Western Branch of the Russian Academy of Medical Science, 3, Dynamo Prospect, St-Petersburg, 197110, Russia.
[2] Institute of Macromolecular Compounds, Russian Academy of Sciences, 31, Bolshoy Pr., St.Petersburg, 199004, Russia.
[3] Centre for Mathematical Biology, Mathematical Institute, University of Oxford, 24–29 St Giles', Oxford 0X1 3LB, UK.

Correspondence to: Prof. Vladimir Khavinson,
St.Petersburg Institute of Bioregulation and Gerontology
3, Dynamo Prospect, St-Petersburg 197110, RUSSIA
TEL + 7 (812) 235-1832, 230 0049
FAX + 7 (812) 235-1832, 230-0049
EMAIL khavinson@gerontology.ru

Submitted: April 27, 2004 *Accepted: August 3, 2004*

Key words: **epitalon; geroprotector; gene promoter; base pair sequence; regulatory peptides; double-helix interface; complementary interactions**

Neuroendocrinol Lett 2005; **26**(3):237–241 **PMID:** 15990728 NEL260305A08 ©Neuroendocrinology Letters **www.nel.edu**

Abstract

PROBLEM. Gerontology observations show that ageing of organism is accompanied by the decrease of the chromatin activity and slowing down of the protein synthesis. Natural regulatory oligopeptides and their synthetic analogues take part in the activation of chromatin and normalise rate of the protein synthesis in cultured tissues. Regulatory peptides significantly enhance longevity.
OBJECTIVES. The objective of this work is to find a possible molecular mechanism by which the regulatory peptides influence the part of genetic system is responsible for initiation of protein synthesis in higher organisms.
METHODS. An interdisciplinary approach was used to address the problem. The work involved not only medical scientists and biologists, but also specialists in biopolymer chemistry, and mathematicians for statistical analysis of information interaction of amino acid and nucleotide sequences.
CONCLUSIONS. The structures and metrics of peptides and the DNA double-helix cause the recognition and complementary binding of a regulatory peptide with DNA functional groups at the interface of the major groove. We have used complementary binding model to find a possible base pair sequence ATTTTC for specific binding of synthetic tetrapeptide epitalon. This base pair block and its reverse complement were found repeatedly in the promoter region of telomerase.
A comparison of statistical information content of peptides and oligonucleotide sequences shows that these two classes of biopolymers are information carriers and. exchange the information to initiate the gene transcription.

Introduction

Recent experiments show that natural peptide preparations and synthetic analogues - regulatory peptides (RP) are involved in the chromatin activation, normalise rates of protein synthesis and contribute to lifespan increase [1–4].

The transcription activation in eukaryotes requires tens of macromolecular activators, transcription factors (TF), and peptides [5,6]. The study of the detailed interactions between proteins and non-coding DNA structures responsible for transcription initiation evolved over the past decade to the understanding that recognition arises when oligopeptide of definite amino acid sequence interacts with preferable base pair sequence of DNA double helix [7 - 9]. However, currently there exists an obvious gap between the numerous evidences of specific effects caused by RP in the activation of protein synthesis and limited modelling of molecular mechanisms which underlie the selective peptide binding with specific DNA site.

The non-coding parts of DNA contain multiple repeats of relatively short base pair sequences (nucleotide blocks). The typical length of these blocks within promoter regions is 6–10 bp, so that they do not exceed one helix turn [10]. Macromolecular TFs interacts with cognate DNA elements exactly of that size. For example the octamer motif ATTTGCAT can be found in the promoters of various ubiquitous genes and binds B-cell-specific TF [11]; every subunit of macromolecular activator protein (CAP) interacts with 8 bp.DNA- site [9].

Regulatory peptides and TFs also contain repeats of characteristic amino acid sequences amino acid blocks. They are normally compiled of 5–6 amino acid residues. We suppose that short peptides, analogous to these amino acid blocks, may serve as agonists for TF, and the initial start signal for site-specific binding of the RP in the DNA major groove.

Our model for sequence specific interaction of RP with recognition site in the promoter region is built on the principles of geometrical and electrochemical complementarity of two right-handed helices: peptide and DNA. The highest selectivity of recognition may be achieved when amino acid block interacts with nucleotide block complementarily. The literature data on molecular geometry of the DNA double-helix and peptide chain were used to find a possible base pair sequence for specific binding of synthetic tetrapeptide epitalon [5, 10].

Foundation principles of complementary interactions in the system "RP - DNA"

Molecular recognition in biology is based upon precise correspondence between molecular surfaces of interacting molecules and equality of distance parameters, which determine the position of interacting functional groups. Structural correspondence of this kind is usually referred to as matrix complemetarity. A regulatory peptide recognises a specific nucleotide sequence in DNA if its own amino acid sequence is complementary to the DNA sequence to a sufficient extent.

A polypeptide chain consisting of L-isomers can have several conformations. β Sheets are the most stretched and exposed type of peptide conformations, with the average distance 3.47 Å. per one amino acid residue. This conformation allows sufficient freedom for the side groups and let them be involved in the largest number of intermolecular interactions [5]. Orientation of the amino acid side groups depends on their interactions with the environment.

Regulatory (non-coding) regions of the DNA contain sites of selective binding of TF with RPs. There are ten base pairs within a single turn of relaxed DNA double-helix, and base pairs are situated almost in parallel planes. Average distance between the planes is 3.4 Å [10]. It is known that phosphodiester carcass chains divide the cross section of DNA double-helix into two unequal parts: the major and the minor grooves, with widths of 22.2 Å and 11.8 Å, correspondingly [10].

Thus, a certain metric correspondence between the DNA double-helix and the peptide chain can be observed: the length is 3.47 Å per one amino acid residue in β-conformation of a peptide chain, while the distance between DNA base pairs is 3.4 Å. As both molecules are right-handed helices, the oligopeptide can easily coincide with the major groove and interact on the interface with functional groups of base pairs. Similar polyfunctional interaction of DNA with ligands was studied in detail and described for an antibiotic netropsin as it forms complex with the minor groove of the double-helix [12,13]. Netropsin has three hydrophobic sites and a positive charge at the end of the molecule. We extended this approach to the case of oligopeptide interaction with functional groups of the DNA major groove, where side groups of RP can be inserted.

The model of complementary interaction of tetrapeptide AEDG with DNA double-helix

Here we present a model of site-specific interaction of DNA with tetrapeptide *AEDG** (epitalon) which is known to have repair properties which cause decrease in the number of tumours and increase in the average longevity [1–4].

Investigation of the tetrapeptide conformational structures showed that the mostly energetically beneficial conformation in water (uniform environment) is of 15.5 Å in length and 8.5 Å in width (maximum). In the absence of low molecular weight electrolyte the negative charge of the C-terminal carboxylic group is compensated by protonated N-terminal amino group, i.e. intramolecular interaction partially cycles the tetrapeptide. Within salt solution and at polyfunctional interaction with DNA the tetrapeptide takes a stretched shape presented in Figure 1. Obviously, under the physiological conditions N-terminal amino group may participate in H-bonds and negatively charged carboxylic groups are involved into polar and ion-ion interactions. The side methyl group

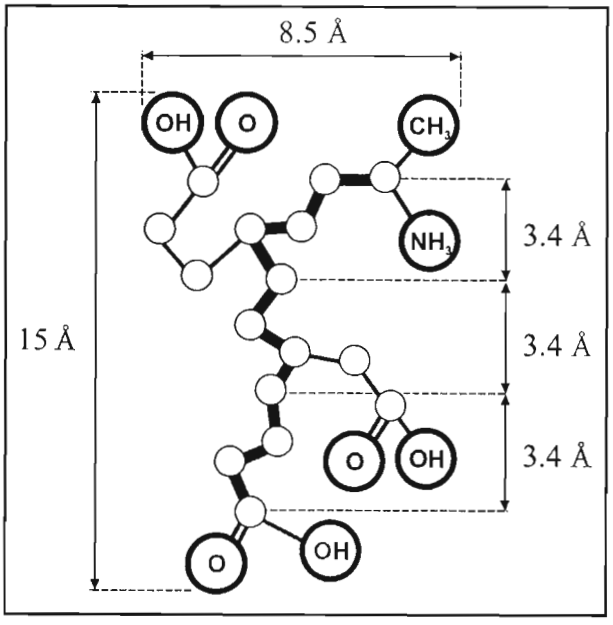

Figure 1. Stretched conformation of peptide Ala-Glu-Asp-Gly (projection to plane). Those terminal and side groups which are capable of complementary interaction with DNA are shown.
$-NH_2$, -OH – proton donating groups;
= O – proton accpeting groups; :
– CH_3 – hydrophobic (methyl) group
The main peptide chain is shown with the thicker line.

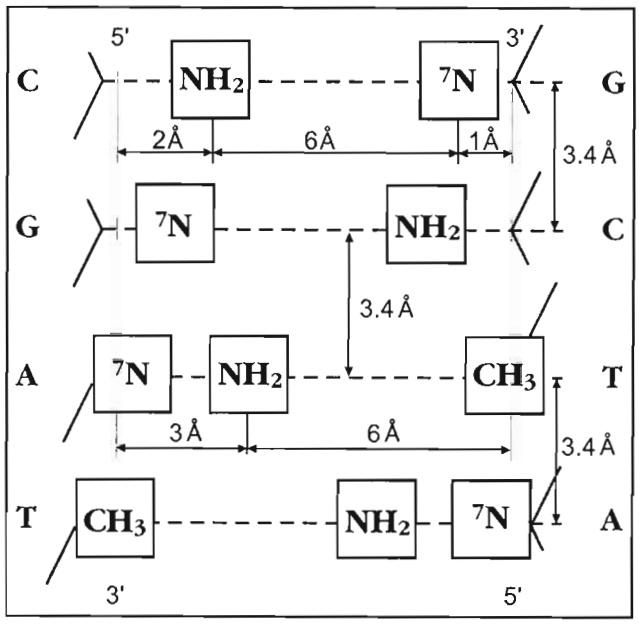

Figure 2. Metric position of functional groups on the surface of the major groove when each base pair is inserted into the DNA double-helix. Broken line shows the perpendicular plane containing aromatic structures of the nucleotide bases.
$-NH_2$, – proton donating groups;
= 7N – proton accpeting groups;
CH_3 – hydrophobic (methyl) group

of alanine residue can interact with hydrophobic groups of nucleic bases. The net electrostatic charge of the peptide is negative; thus, there are no interactions with phosphate groups of phosphodiester carcass chains.

We suppose that the tetrapeptide with rather capacious side groups can be placed only inside the major groove and can interact with those functional groups of the nucleic bases which are exposed on the surface of the major groove of the double-helix. Metrical characteristics of nucleotide functional groups exposed onto the surface of the DNA major groove are shown in Figure 2.

Each sequence of base pairs produces a unique pattern of the exposed functional groups. They can be included into H-bonds or ion bonds and into hydrophobic interactions with peptide side groups if their sequence complementary corresponds to this pattern. Polar interactions of RP with base pairs in the major groove determined by electrostatic interactions of carboxylic groups of the peptide with amino groups of adenin and cytosine, and also by H-bond of the protonised amino groups and carboxylic groups of the peptide with the nitrogen atom N^7 of adenin and guanin [9]. Hydrophobic interactions in this system are due to the side groups of the peptide and methyl groups of thymine

.A simple permutation of nucleotide pairs (functional groups are shown in Figure 2) allows to find their sequence that provides complementarity with side groups of the peptide. The complementary nucleotide sequence for the tetrapeptide AEDG is ATTTC (Figure 3). A distinctive feature of our model is the simultaneous interaction of the peptide with functional groups of both chains of DNA double-helix. To find geometric complementarity we take into account lengths of covalent bonds 1.5 Å, ion-ion bonds 2.5 Å and H-bonds 2.0–2.5 Å. We consider impossible peptide interactions with phosphate groups of DNA carcass chains since these interactions are not specific. We did not consider possible changes of the angles between base pair planes and the axis of the double-helix either.

Intermolecular interactions between AEDG peptide and DNA are based on several H-bonds. For peptide systems the length of H-bond is, typically, 2.0–2.5 Å, i.e. is larger than a half of the distance between the neighbouring base pairs planes.

The complex of the tetrapeptide AEDG with DNA is shown in Figure 4. The tetrapeptide *AEDG* is located inside the major groove of the DNA according to complementarity of the peptide and DNA functional groups. An important point to notice is that due to high mobility of functional groups of terminal glycine residue, the tetrapeptide AEDG can complementary bind also with ATTTG sequence of DNA: although the bonds will be weaker, than for ATTTC.

Ageing of the organism is known to be accompanied by slowing down protein synthesis. Transcription of all protein-coding genes is performed by the enzyme DNA-dependent RNA polymerase II after TF have initiated the process [14]. Biosynthesis of RNA polymerase II also cannot be done unless the transcription have been initiated by a corresponding TF.

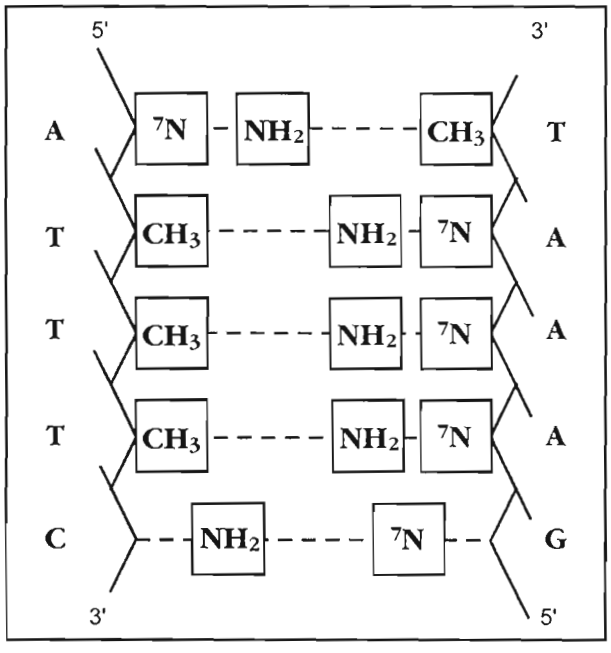

Figure 3. Sequence of nucleotide pairs in the DNA double-helix with functional groups that are complementary to the functional groups of the peptide Ala-Glu-Asp-Gly. There are mulitiple repeats of this sequence of base pairs in the promoter of telomerase gene.

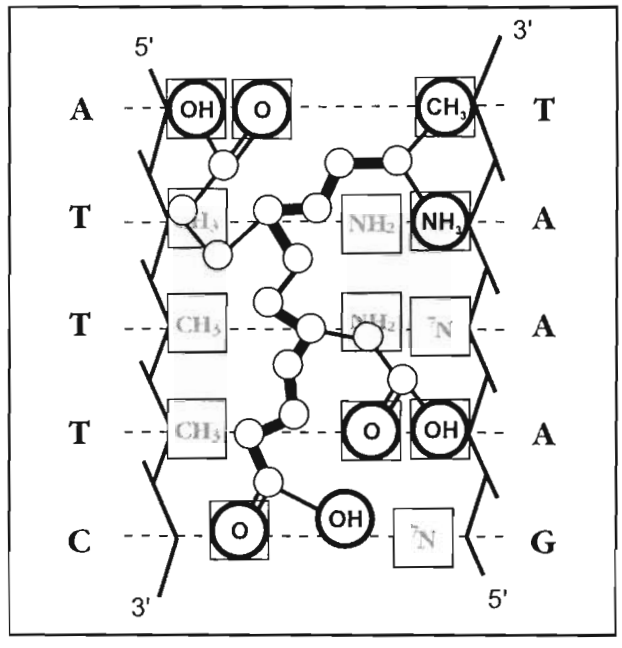

Figure 4. Model of complementary interaction of the peptide Ala-Glu-Asp-Gly with DNA double-helix (DNA-peptide complex on the promoter region of telomerase gene).

It had been reported that the block ATTTGCAT is one of the binding sites for the RNA polymerase II TF[11]. Within promoter of the large subunit RNA polymerase the sequence ATTTC is present three times [15].

As it was mentioned above, the tetrapeptide AEDG causes a special interest exactly due to its high geroprotective efficiency. It was shown in animal tests that it significantly increases longevity and slows down processes of tumour formation. There is a hypothesis which links the general lifetime of cell populations to the length of chromosome telomers, and so, to the activity of the enzyme telomerase [16]. From the literature data we found that the nucleotide sequence ATTTC/G, which is the predicted DNA sequence from our model, is present nine times in the promoter region of telomerase protein component (in the segment 3729 b.p.from the start of transcription) [17]. The complementary interaction of the oligopeptide AEDG with the blocks ATTTC seem to cause reactivation of telomerase promoter in the somatic cells which initiates intracellular synthesis of telomerase and elongation of telomers, thus enhancing proliferating potential of tissue and may considerably affect longevity. The synthetic tetrapeptide AEDG then can be believed to exhibit geroprotective function as a TF agonist, interacting with the nucleotide blocks in promoter regions of vitally important genes, i.g. RNA polymerase and telomerase.

Informational interaction in the system "RP — DNA"

The complementary interaction of the peptide chain with DNA not only enlarges the major groove of the double-helix, but also increases the local rigidity of these biopolymers. A corresponding descrease in the conformational entropy for both peptide and DNA can be considered as an informational signal. In order to obtain a quantitative description of the informational signal we estimate information charge of the peptide and DNA. In other words, we calculate how much information in bits contains single structural element (letter) in the sequences where elements are amino acid residues and nucleotides.

Informational content of the short repeated blocks within the long sequences of amino acid residues or nucleotides was calculated as proposed by Shannon [18] using a software available at www.maths.ox.ac.uk/'chernova/codes/belok.html

We searched for repeated amino acid blocks within sequences of regulatory proteins and peptides and found values for the information charge using N-gramm Shannon entropies [19]. The results of calculations which were done by [20] show that the most often repeated blocks in polypeptide chains consist of four-five amino acid residues and carry a certain information charge which reaches values up to 8 bit per residue within block [21]. Repeated blocks in peptide chains which carry information charge are the most evolutionary conserved amino acid sequences in eukaryota. Obviously that signal molecules must be compact and short – otherwise they would easily

loose the accuracy of signalling function in the course of mutational changes.

Repeated blocks and information charges were calculated in the same way for the nucleotide sequences of gene promoters of RNA polymerase II [15] and telomerase [17]. We found that the non-coding regions usually contain repeated blocks of seven nucleotides, for which the value of information charge is up to 6.4 bit per base pair.

The number of symbols in the "alphabets" for amino acids and nucleotides is different, twenty and four, correspondingly. In order to transmit (translate) information from donor to accepter (from one language to another) without loses and distortions it is necessary to require that the donor and accepter have close information capacity. In the case of "RP-DNA" interaction the requirement is to have repeated blocks made up of five to seven nucleotides.

Molecular model of complementary interaction of the regulatory tetrapeptide with the DNA double-helix can be supplied with the concept of discrete, rather quantum, information transmission during the molecular contact. The media for information transmission are short blocks (oligomers) which are preserved within macromolecular biopolymers, (proteins and nucleic acids), over millions years of evolution.

Consideration of complementary interactions between oligopeptides and oligonucleotides is an important part in the core of modern hypotheses on the initial stages of living matter, especially from the point of view of genesis and accumulation of molecular information during evolution. The combination of permanently acting mechanisms of mutations and selection with a stable and mechanism of conservation of molecular information determines the direction and the irreversibility of evolutionary processes.

REFERENCES

1 Khavinson VKh. Peptides and Ageing. Neuroendocrinology Letters 2003; **23** Suppl. 3. pp. 144 p.

2 Khavinson VKh, Izmailov DM, Obuchova LK, Malinin VV Effect of epitalon on the lifespan increase in *Drosophilla melanogaster*. Mech Ageing Dev 2000; **120**:141–149.

3 Khavinson V, Goncharova N, & Lapin B. Synthetic tetrapeptide epitalon restores disturbed neuroendocrine regulation in senescent monkeys. Neuroendocrinology Letters 2001; **22**:251–254.

4 Anisimov VN, Khavinson VKh, Popovich IG, Zabezhinski MA.Inhibitory effect of peptide Epitalon in colon carcinogenesis induced by 1,2-dimethylhydrazine in rats. Cancer Letters 2002; **183**:1–8.

5 Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Molecular cell biology. /Library of Congress cataloging-in-publication data; 2000.

6 Woychik NA, Hampsey M. The RNA polymerase II machinery: structure illuminates function. Cell 2002; **108**:453 - 463

7 Parkinson G, Wilson C, Gunasekera A, Ebright YW, Ebright RH, Berman HM. Structure of the CAP - DNA complex at 2.5 Å resolution: a complete picture of the protein-DNA interface. J Mol Biol 1996; **260**:395–408.

8 Parkinson G, Gunasekera A, Vojtechovsky J, Zhang X, Kunkel TA, Berman H, Ebright RH. Aromatic hydrogen bond in sequence-specific protein DNA recognition. Nat Struct Biol 1996; **3**:837 - 841.

9 Lawson CL, Swigon D, Murakami KS, Darst SA,Berman HM, and Ebright RH. Catabolite activator protein: DNA Binding and transcription activation. Current opinion in structural biology 2004; **14**:10–20.

10 Van Holde KE. Chromatin.NY: Springer-Verlag; 1988.

11 Takechi S, Adachi M, Nakayama T. Cloning and characterization of the chick JCT binding factor OBF-1. Biochem Biophys Acta 2002; **1577**:466 - 470.

12 Gallagher K, Sharp K. Electrostatic contributions to heat capacity changes of DNA-ligand binding. Biophys J 1998; **75:** 769 -776.

13 Rentzeperis D, Marky LA. Interaction of minor groove ligands to an AAATT / AATTT site correlation of thermodynamic characterization and solution structure. Biochemistry 1995; **34:**2937 -2945.

14 Dynlacht BD Regulation of transcription by protein, that control the cell cycle. Nature1997;. **389**:149–152

15 Mita K, Tsuji H, Morimyo M, Takahashi E, Nenoi M, Ichimura S, Yamauchi M, Hongo E, Hayashi A. The human gene encoding the largest subunit of RNA polymerase II.Gene1995; **159**:285–286.

16 Blasco MA. Telomerase beyond telomeres.Nature Reviews 2002; August:1– 6

17 Wick V, Zubov D, Hagen G. Genomic organisation and promoter characterization of the gene encoding the human telomerase reverse transcriptase (hTERT). Gene 1999; **232:**97–106

18 Shannon C. Prediction and entropy of printed English. // Bell System Tech J,1951; **30:**50–64

19 Atchley WR, Wollenberg KR, Fitch WM, Terhalle W, Dress AW. Correlation among Amino Acid Sites in bHLH protein domains: an information theoretic analysis. Molec Biol Evolution 2000; **17**: 164 -178.

20 Chernova AA. http://www.maths.ox.ac.uk./~chernova/codes/belok.htm

21 Shataeva LK, Chernova AA, Khavinson VKh. Information charge of amino acid sequences.J Biol Chem, in press