Short Cell-Penetrating Peptides: A Model of Interactions with Gene Promoter Sites

V. Kh. Khavinson[1,2], S. I. Tarnovskaya[2], N. S. Linkova[2], V. E. Pronyaeva[2], L. K. Shataeva[2], and P. P. Yakutseni[3]

Translated from *Byulleten' Eksperimental'noi Biologii i Meditsiny*, Vol. 154, No. 9, pp. 391-396, September, 2012
Original article submitted January 10, 2011

Analysis of the main parameters of molecular mechanics (number of hydrogen bonds, hydrophobic and electrostatic interactions, DNA-peptide complex minimization energy) provided the data to validate the previously proposed qualitative models of peptide–DNA interactions and to evaluate their quantitative characteristics. Based on these estimations, a three-dimensional model of Lys-Glu and Ala-Glu-Asp-Gly peptide interactions with DNA sites (GCAG and ATTTC) located in the promoter zones of genes encoding CD5, IL-2, MMP2, and Tram1 signal molecules.

Key Words: *cell penetrating peptides; molecular simulation; DNA-peptide interactions*

Aging of an organism correlates with inhibition of some metabolic processes, which is explained by lesser intensity of gene transcription and mRNA synthesis [1]. Gene expression in eukaryotic cells is triggered by transcription factors, high molecular proteins and low molecular peptides [5,11].

Short cell-penetrating peptides (CPPs) are involved in activation of cell proliferation and differentiation, realized by transcription factors. Cell penetrating short peptides are a group of peptides consisting of no more than 20 amino acid residues with molecular weight of no more than 4 kDa [3]. Normally CPPs are multicharge ions. They noncovalently bind to nucleic acids, amino acids, peptides and transport them to the site of destination inside the cell [3]. The CPPs group includes natural (R-PTD_4, bt-NLS, Lig1-PBD-F, F(Ahx)-TAT), and short synthetic peptides [2,13], for example, Lys-Glu (vilone) and Ala-Glu-Asp-Gly (epithalone), created at St. Petersburg Institute of Bioregulation and Gerontology. The mechanism of penetration of highly hydrophilic CPPs into the cell and into the genome remains unclear. The relevant hypotheses (direct penetration through membranes, endocytosis, micellar internalization) remain an object of discussions [14]. In contrast to steroid hormones, hydrophilic short peptides bind to phospholipid hydrophilic groups on the outer side of the plasma membrane, group together, and enter the cell by the mechanism close to pinocytosis [4,14].

A specific feature of the nuclear membrane is a well-developed system of transport pores formed by protein complexes, nucleoporines that control the transport of nucleoprotein complexes into and from the nucleus [10]. The inner diameter of nucleopores is 42 nm, outer diameter 50 nm. Hence, they are permeable for diffusing molecules with molecular weight of up to 5 kDa [9] and hence, for Lys-Glu and Ala-Glu-Asp-Gly peptides. Peptides Ala-Glu-Asp-Gly, Glu-Asp-Arg, and Lys-Glu-Asp-Gly labeled with FITC really penetrate into the cytoplasm, nucleus, and nucleolus of HeLa cells [4]. The mechanism of their penetration into the nucleus may be similar to that described for natural CPPs, but it is also possible that Lys-Glu and Ala-Glu-Asp-Gly are transported to the nucleus by larger natural CPPs.

It is known that Lys-Glu and Ala-Glu-Asp-Gly CPPs regulate the expression of genes functionally belonging to different cell systems [6]. Hence, they

[1]I. P. Pavlov Institute of Physiology, Russian Academy of Sciences, St. Petersburg; [2]St. Petersburg Institute of Bioregulation and Gerontology; [3]St. Petersburg State Polytechnical University, Russia. ***Address for correspondence:*** miayy@yandex.ru. N. S. Linkova

act selectively towards the binding sites on the gene promoter sites. It is assumed that the peptide amino acid residues form a network of hydrogen bonds with the functional groups in the DNA double spiral greater groove [12].

The Ala-Glu-Asp-Gly peptide reduces the risk of tumor development, regulates the functional activity of the pineal gland, pancreas, and thymus, promotes elongation of telomeres, and prolongs the mean lifespan [6]. The Lys-Glu peptide stimulates immunogenesis in the thymus, spleen, and bone marrow, is characterized by radioprotective, geroprotective, and immunomodulating effects [2].

We have previously proposed qualitative models of complementary binding of Ala-Glu-Asp-Gly peptides to ATTTC, GTTTC, and CTTTC sequences of the gene promoter site and of Lys-Glu binding to CGAG sequence [7].

Now we carried out three-dimensional computer simulation of the processes in order to evaluate quantitatively the interactions between CPPs (Lys-Gly, Ala-Glu-Asp-Gly) and DNA sites (ATTTC and GCAG) in water medium.

MATERIALS AND METHODS

The main method for estimations was the molecular mechanical method with different fields of forces (MM^+, Amber99, Opls, Charmm27). This method was selected due to its high productivity: the molecule atoms could be regarded as the Newton particles interacting with each other by preset field of forces and hence, it was possible to sufficiently precisely analyze all probable variants of spatial structure of the main chains of the studied peptides and of the corresponding submultitudes of the labile lateral chain conformations. The molecular dynamic procedures, reproducing the movement of isolated molecules in a certain time interval, were used for preliminary selection of the main energetically the most beneficial conformations of Ala-Glu-Asp-Gly and Lys-Glu. The results were then verified by optimization of the geometry in different fields of forces. The main results of molecular dynamics were obtained in the MM^+ field of forces in two modes: (1) standard, 1 psec long, with a 0.001 psec step and 310°K temperature of the system and (2) by simulated annealing with 1000°K for the initial temperature of the system, cooling to 100°K, and to 0°K after 200 steps. The geometry optimization was then carried out for pre-selected conformations (preset minimization gradient 0.05 kcal/Å×mol^{-1}) and the precise energy of the resultant conformations was calculated. Speaking about the minimization energy, we mean energetically the most beneficial status of the molecule. About 100 preliminary peptide conformations were analyzed in each mode. Energetically the most beneficial rotameres were then selected from them.

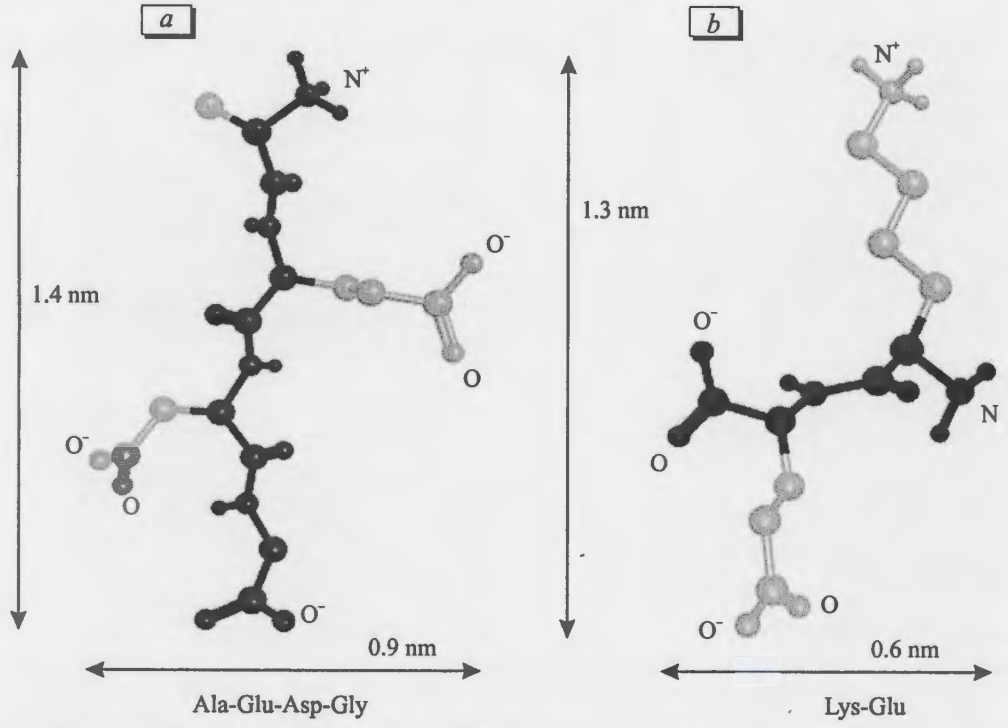

Fig. 1. Conformations of Ala-Glu-Asp-Gly (*a*) and Lys-Glu (*b*) with optimum minimization energy. The main chain of the peptide including carboxyl and amino groups is shown in black and the lateral chains are gray.

The ATTTC and GCAG gene sites were simulated using the DNA structures from PDB database: 1A02, 1QMS. The impact of the solvent in estimation of the DNA–peptide complex minimization energy was taken account of in the generalized Born's approximation with introduction of the internal dielectrical constant, equal to 1, and external one, equal to 80. Estimations of the complexes were carried out in the AMBER99 field of forces. The energy of peptide binding to DNA (E, kcal/mol) was calculated as the difference of individual molecules of DNA and peptide and the DNA-peptide complex.

RESULTS

Energetically the most beneficial conformations of Ala-Glu-Asp-Gly and Lys-Glu peptides were obtained by the molecular mechanical method (Fig. 1). These peptide designs have a β-structure and belong to the peptide unfolded conformation shape, providing the maximum steric freedom of the lateral groups of peptide chain, allowing realization of the maximum number of interactions between molecules. Normally (T=310°K, pH 7) the terminal groups may be zwitter ions, the peptides turning into charged molecules due to the lateral carboxyl and amino groups. Evaluation of deprotonation degree has shown that a summary -1 charge can emerge on Lys-Glu and -2 charge on Ala-Glu-Asp-Gly. In addition, the Ala-Glu-Asp-Gly molecule is formally less hydrophilic than Lys-Glu.

Transport of substances is determined by the combination of their steric and physicochemical characteristics. The Ala-Glu-Asp-Gly and Lys-Glu molecules differ little by length and are significantly smaller than the nucleopores (Fig. 1). The difference in diameters is greater, and we may expect that Lys-Glu would easier penetrate through the lipid fraction of cellular or nuclear membranes.

The situation seems less obvious when we analyze the physicochemical properties. First, the process is insufficiently well studied: the mechanisms of CPPs penetration remain unclear. Second, hydrophilicity. During common passive diffusion less hydrophilic substances easier diffuse through the lipid bilayer and it seems that Ala-Glu-Asp-Gly would have some advantages. High hydrophilicity of Lys-Glu is determined by the presence of lysine residue; however, this is also an amphiphilic fragment capable of stimulating, if not transport, then preferable associations of the dipeptide with the membrane. Third, the environmental effects. In a real medium, the local microenvironment of the biphase is heterogeneous and changes to the degree equivalent to the distance between atoms. As a result, a molecule or even part of it can transform from charged to electrically neutral form or vice versa. Hence, the prognosis for transport in the lipid membrane or nucleopore is liable to change and is therefore a problem beyond the tasks of this study.

Calculations of the interactions of the studied peptides with the DNA molecule were performed for right-handed B-form DNA molecule. A large and a

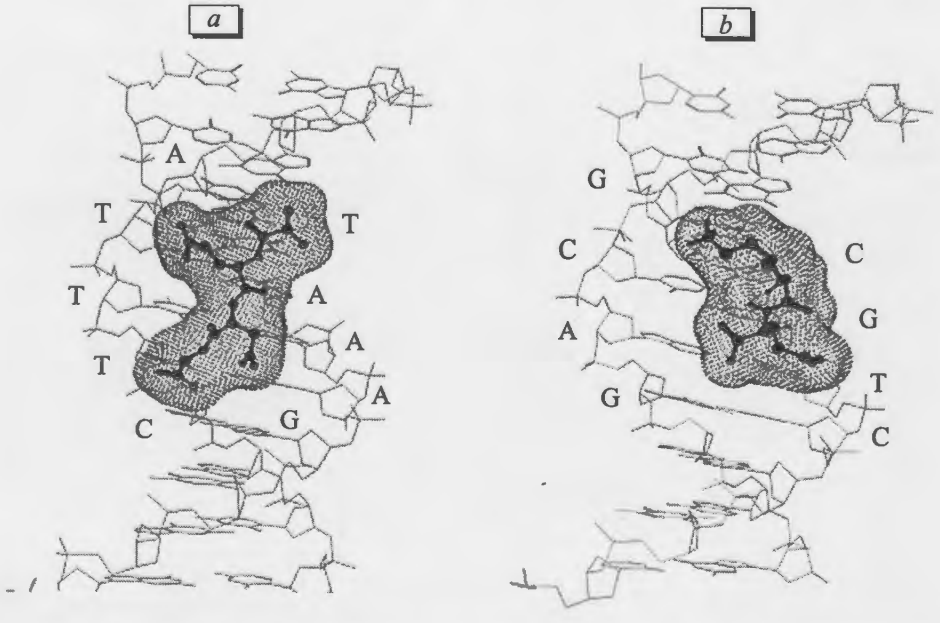

Fig. 2. Location of peptides in putative binding centers of gene promoter regions. *a*) Ala-Glu-Asp-Gly location in IL-2 gene site; *b*) Lys-Glu location in CD5 gene site. Dotted line shows the peptide molecular membranes, Here and in Fig. 3: A: adenine; T: thymine; G: guanine; C: cytosine.

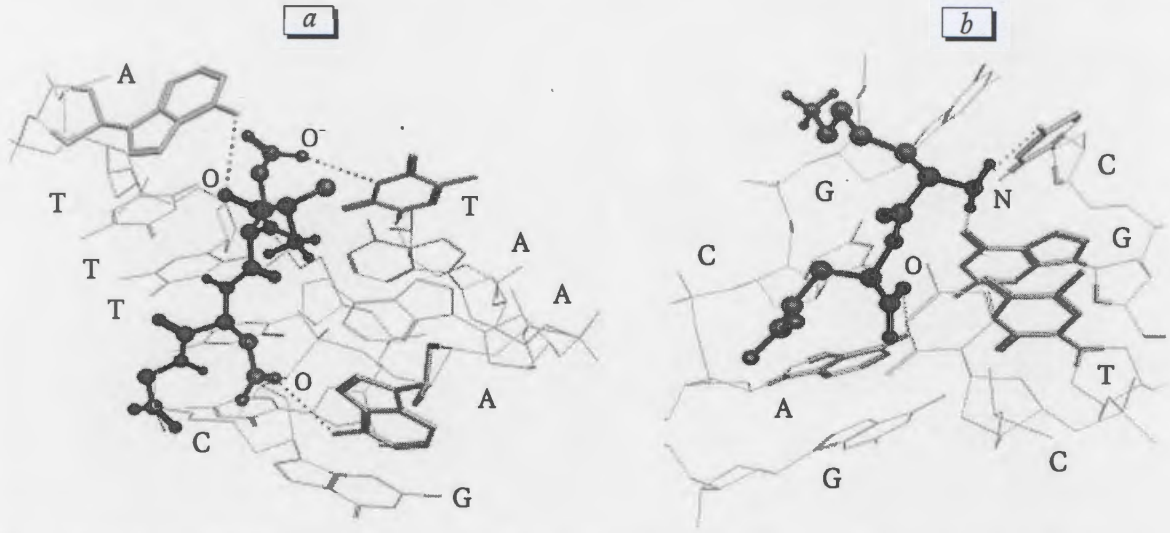

Fig. 3. Peptide interactions with DNA nitrous bases. *a*) Ala-Glu-Asp-Gly interaction with ATTTC sequence; *b*) Lys-Glu interactions with GCAG sequence. Dotted line shows hydrogen bonds between peptide and DNA atoms, bold line show DNA nitrous bases, forming a hydrogen network with the peptide. N: nitrogen; O: oxygen.

small groove, 2.1 and 1.2 nm wide, respectively, could be discerned on the surface of the double strand molecule. The Ala-Glu-Asp-Gly and Lys-Glu interacted with DNA by binding to its large groove (Fig. 2).

The position of the bound peptides was discretely fixed. Free continuous movements along the grooves were impossible. A peptide chain step in the β-conformation per amino acid was equal to 0.35 nm, the distance between base pairs in the DNA chain being 0.34 nm. DNA regulatory sites (Fig. 2) contain sites for selective binding to the studied peptides (Table 1) and the detected models of Ala-Glu-Asp-Gly interactions with ATTTC binding site and of Lys-Glu interactions with GCAG binding site (Fig. 3).

Interactions of Ala-Glu-Asp-Gly with ATTTC sequence and with TAAAG complementary to it was realized via the van der Waals, electrostatic interactions and formation of hydrogen bonds between the functional groups of both molecules. The complex formation energy was 10.3 kcal/mol. The network of three hydrogen bonds was found, formed by oxygen atoms in alanine main chain and adenine nitrogen-containing fragment, glutamic acid lateral chain carboxyl group and thymine nitrogen-containing fragment, and aspartic acid lateral chain carboxyl group and adenine nitrogen-containing fragment (Fig. 3, *a*). Interactions of Lys-Glu with GCAG sequence and the complementary CGTC were realized in the presence of three hydrogen bonds: the lysine main chain amino group formed hydrogen bonds with guanine oxygen atom and cytosine nitrogen, while glutamic acid main chain carboxyl group formed a hydrogen bond with adenine nitrogen atom (Fig. 3, *b*). This peptide binding energy was equal to 11 kcal/mol. Hence, Lys-Glu served as an extra suture in the gene promoter site. The Lys-Glu and Ala-Glu-Asp-Gly interaction energies were similar: 10.3 and 11.0 kcal/mol, respectively.

Similar values of peptide binding energy may indicate similar pharmacodynamic activities of the studied substances. The absolute values of energy of the ligand-DNA interactions, obtained in the study, corresponded to energies typical of substances with pronounced biological effects towards their targets.

The designed models of Lys-Glu and Ala-Glu-Asp-Gly interactions are an important aspect for understanding their interactions with some genes and the synthesis of the respective proteins (Table 1).

The GCAG sequence, with which, according to simulation data, Lys-Glu bound, was detected in the promoter zone of the gene, modulating the CD5 transmembrane protein, marker of immune cell differentiation. A previous study showed that the expression of CD5 molecule on lymphocyte precursors in the pineal gland, thymus, and spleen changed under the effect of Lys-Glu dipeptide, this indicating these cells' proliferation and/or differentiation [2]. The ATTTC sequence, complementary to Ala-Glu-Asp-Gly, was detected in the promoter zones of IL-2, matrix metalloproteinase-2 (MMP2), and TIR domain-containing adaptor molecule 1 (Traml) genes. A previous study showed that Ala-Glu-Asp-Gly tetrapeptide regulated the activities of immune cells, including their production of cytokines, one of which is IL-2 [8].

In addition, Ala-Glu-Asp-Gly restores the expression of MMP2 protein involved in regulation of intercellular matrix in the pineal gland and thymus; Traml is an intracellular adapter protein involved in the functioning of antibacterial defense receptor, while

TABLE 1. Peptide Binding Sites in Gene Promoter Zones

Gene	Location of probable binding sites on gene for peptide	Peptide
CD5 (rat)	1ctccaagagagtaagacactgagccagacacctttcctg gctggccaactgaactcacct 61gaggctgaggcagcagaaggccattatcc	Lys-Glu
IL-2 (human)	1cgaattcccctatcacctaagtgtgggctaatgtaacaa a gagggatttcacctacatcc 61attcagtcagtctttgggggtttaaagaaattccaaag ag tcatcagaagaggaaaaatc 121aaggtaatgtttttttcagacaggtaaagtctttgaaa ata tgtgtaatatgtaaaacatt 181ttgacacccccataatatttttccagaattaacagta taa attgcatctcttgttcaaga 241gttccctatcactctctttaatcactactcacagtaa cct caactcctgccaca	Ala-Glu-Asp-Gly
MMP2 (rat)	1ctggcgtctgcccgcccttgtttccgctgcatccagact t ccccgggtggctggaggctc 61tgtgtgcatccagaactttagatatacaaaggggattac ta ggacctgcaagcacccgcag 121ccgtggtgcgtactggtacgtgggatcccgttatgag acc ctgagcccggagaagctgag 181gcaattgagtaaaggggtctcagaacgccgtggagag cag gcgccagccgggtggacccc 241agggcacagccagcgacctcagggtgacacgcggagc ccg ggagcgcaagg	Ala-Glu-Asp-Gly
Tram 1 (mouse)	1ggatagtcgagtttttctacatctcccagctggcgtact g gctccatgcattccctgaac 61tctacttccagaaaaccaaaaaagaagatatccctcgt cagctcgtctacattggccttt 121atctctttcatattgctggagcctatcttctgaactt gaaccacctgggccttgtccttc 181tggtgctgcactatttcgttgaatttcttttccacat ttctcgtctattttattttagtg 241acgagaagtatcagaaaggggttttctctctgggcagt tctttttgttttgggaagacttc 301tgactctaatttctttcagttcttactggtgggtttgg g	Ala-Glu-Asp-Gly

Note. Putative sites of peptide binding to genes are underlined. Gene promoter zones are taken from GenBank database (NCBI). Figures show the ordinal number of nucleotides in the gene.

the expression of its gene under the effect of the tetrapeptide indicates its immunostimulatory activity.

Our analysis of Ala-Glu-Asp-Gly and Lys-Glu conformations made it possible to create models of site-specific interactions of the peptides with nitrous bases of the regulatory sites of genes encoding CD5, IL-2, MMP2, and Tram1 proteins, involved in cell-cell interactions and in activation of the neuroimmunoendocrine system cells. These peptides can bind by complementary binding to sequences forming the large groove of two-spiral DNA, which is shown by their complementary compatibility by the "key-lock" mode. The structural models, obtained by molecular simulation, and the specific features of Ala-Glu-Asp-Gly and Lys-Glu interactions with nitrous bases of ATTTC and GCAG sequences, confirm the previously suggested hypotheses on the probability of formation and type of peptide complexes with nucleic acids [7]. Hence, these

models can be useful for deciphering the mechanisms of biological activities of CPPs, quantitative evaluation of their putative activity, and directed search for new substances with preset pharmacological effects.

REFERENCES

1. B. Alberts, D. Bray, J. Lewis, *et al.*, *Molecular Biology of the Cell*, 3rd ed., New York (1994).
2. V. N. Anisimov and V. Kh. Khavinson, *Biogerontology*, **11**, No. 2, 139-149 (2010).
3. A. Chung, F. Eudes, and Y. S. Shim, *IUBMB Life*, **62**, No. 3, 183-193 (2010).
4. L. I. Fedoreyeva, I. I. Kireev, V. Kh. Khavinson, and B. F. Vanyushin, *Biochemistry* (Mosc.), **76**, No. 11, 1210-1219 (2011).
5. S. C. Harrison, *Nature*, **353**, 715-719 (1991).
6. V. Kh. Khavinson and V. V. Malinin, *Gerontological Aspects of Genome Peptide Regulation*, Basel (2005).
7. V. Kh. Khavinson, L. Shataeva, and A. Chernova, *Neuro Endocrinol. Lett.*, **26**, No. 3, 237-241 (2005).
8. N. S. Linkova, V. Kh. Khavinson, N. I. Chalisova, *et al.*, *Bull. Exp. Biol. Med.*, **152**, No. 1, 124-127 (2011).
9. H. Lodish, A. Berk, S. L. Zipursky, *et al.*, *Molecular Cell Biology*, New York (2000).
10. M. M. Ohno, M. Fornered, and W. Mattaj, *Cell*, **92**, No. 3, 327-336 (1998).
11. S. Ryu, S. Zhou, A. G. Ladurner, and R. Tijian, *Nature*, **397**, 446-450 (1999).
12. N. C. Seeman, J. M. Rosenberg, and A. Rich, *Proc. Nat. Acad. Sci. USA*, **73**, No. 3, 804-808 (1976).
13. G. Tünnemann, R. M. Martin, S. Haupt, *et al.*, *FASEB J.*. **20**, No. 11, 1775-1784 (2006).
14. S. Trabulo, A. L. Cardoso, A. M. Cardoso, *et al.*, *Methods Enzymol.*, **509**, 277-300 (2012).