ISSN 1407-7353 MATERIĀLZINĀTNE UN LIETIŠĶĀ ĶĪMIJA / MATERIAL SCIENCE AND APPLIED CHEMISTRY 2008-16

DETERMINATION OF THE 3D STRUCTURE OF DNA HEXAMERS (ATGCAT, ATCGAT, TAGCTA, TACGTA) BY 2D NMR SPECTROSCOPY

DNS HEKSAMĒRU (ATGCAT, ATCGAT, TAGCTA, TACGTA) 3D STRUKTŪRAS NOTEIKŠANA, IZMANTOJOT 2D KMR SPEKTROSKOPIJU

Kirils Zinovjevs, *student*
RTU, Faculty of Material science and applied chemistry
Latvian Institute of Organic Synthesis
Address: Aizkraukles 21, LV 1006, Riga, Latvia
Phone: 7555985

Kristaps Jaudzems, *assistant, B. chem.*
RTU, Faculty of Material science and applied chemistry
Latvian Institute of Organic Synthesis
Address: Aizkraukles 21, LV 1006, Riga, Latvia
Phone: 7555985

Edvards Liepiņš, *senior researcher, Dr.habil.chem*
Latvian Institute of Organic Synthesis
Address: Aizkraukles 21, LV 1006, Riga, Latvia
Phone: 7555985

Atslēgas vārdi: NMR spectroscopy, DNA, oligonucleotides, stereochemistry

Abstract

The solution structures of self-complementary DNA hexamers with ApT or TpA base steps on each end and with CpG or GpC base steps in the center resembling the biologically important E-box sequences were determined. All 4 hexamers form the antiparallel-stranded right-handed duplexes. The $_1$H-NMR spectra acquired at higher temperatures (up to 25° C) showed that all duplexes are stable. The base-pair step parameters were calculated for four middle steps of all duplexes. The analysis of macroscopic curvature gave order TACGAT>ATGCAT>ATCGAT>TAGCTA . It was proposed that the increased macroscopic curvature of sequence already in uncomplexed state allows for easier energy – saved interactions with proteins.

Introduction

The right-handed antiparallel double-stranded DNA is one of the most essential structures in modern biology. The knowledge of DNA structure is essential for forming complexes DNA-protein or DNA-drug, because the helical parameters effect on the ability of protein or drug to interact with DNA by certain sequence. Although the common structure of DNA is well known, there are variations in the helical parameters of real molecules caused by deviations in relative conformations of the local base pairs. Such deviations could be described by parameters that relate complementary base pairs and sequential base-pair steps, examples are shown in Fig. 1. These parameters give all information needed to restore the DNA structure.
Although the structures of DNA were investigated for last 40 years, there is no way to predict the helical parameters from the sequence theoretically. So, the acute problem is to determine how the sequential changes effect on the macroscopical DNA structure. It is worth to mention that the stability of duplexes is directly connected to the particular sequence of DNA too.
ACGT cis-acting DNA sequence elements have been identified in a multitude of plant genes regulated by diverse environmental, physiological, and environmental cues. In vivo transient and transgenic plant expression studies have shown that these ACGT elements are necessary for maximal transcriptional activation [1]. On the other hand the sequence ACGT plays the significant role in cancer evolution [2]. The oncoprotein c-Myc has an important role in cell proliferation, transformation, inhibition of differentiation and apoptosis. These functions most likely result from the transcription factor activity of c-Myc. As a heterodimer with Max, the c Myc protein binds to the E-box sequence (CACGTG). The purified c-Myc/Max complex binds specifically with a high affinity to its natural target, the rat ODC gene, which contains two adjacent, consensus E-boxes. High affinity binding results from the ability of c-Myc/Max dimers to bind cooperatively to these E-boxes.
In paper we determined the spatial structures of four DNA hexamers, having very similar to E-box sequence, to find out how the systematical permutation of nucleotide sequence influences the DNA shape probably responsible for its recognition by proteins.

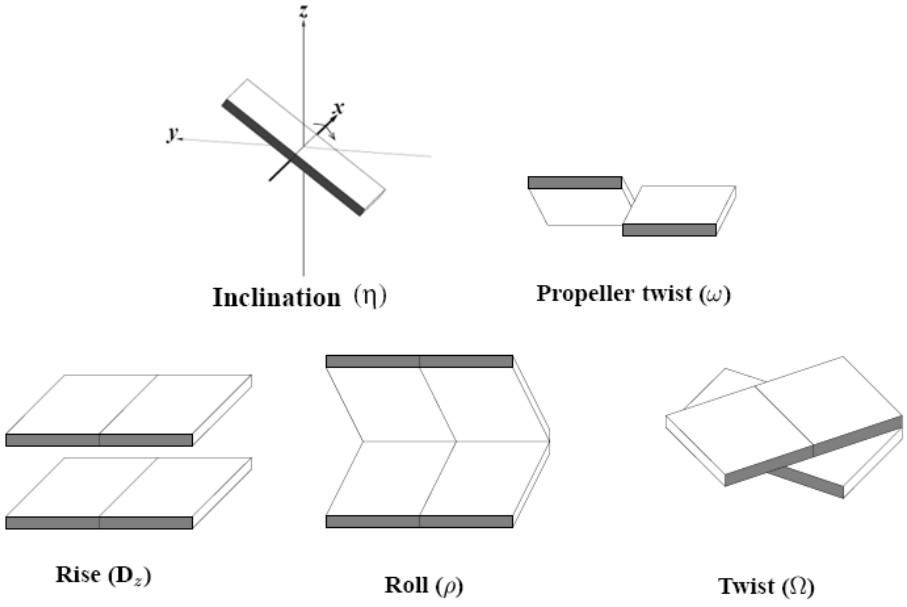

Fig. 1. *Pictorial definitions of parameters that relate complementary base pairs and sequential base-pair steps.*

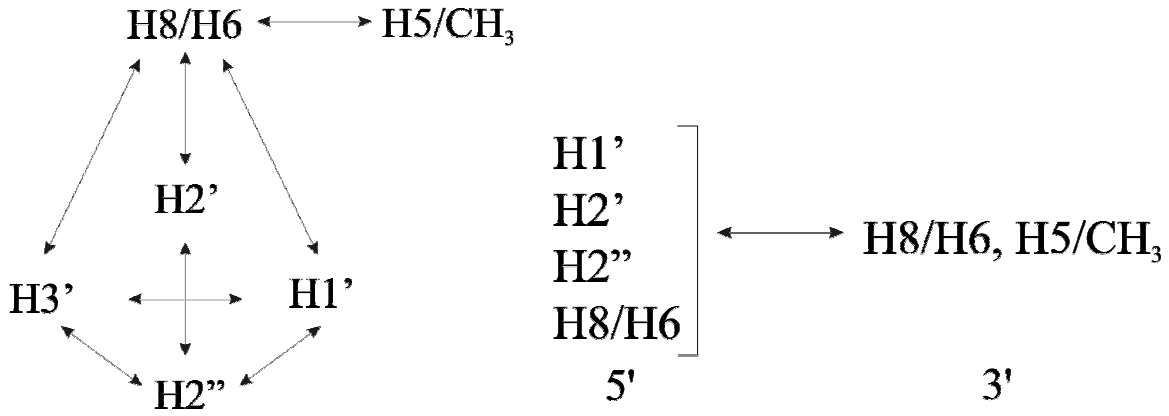

Fig. 2. *Schematic illustration of the intranucleotide and sequential internucleotide interproton distances with values of ≤ 5Å.*

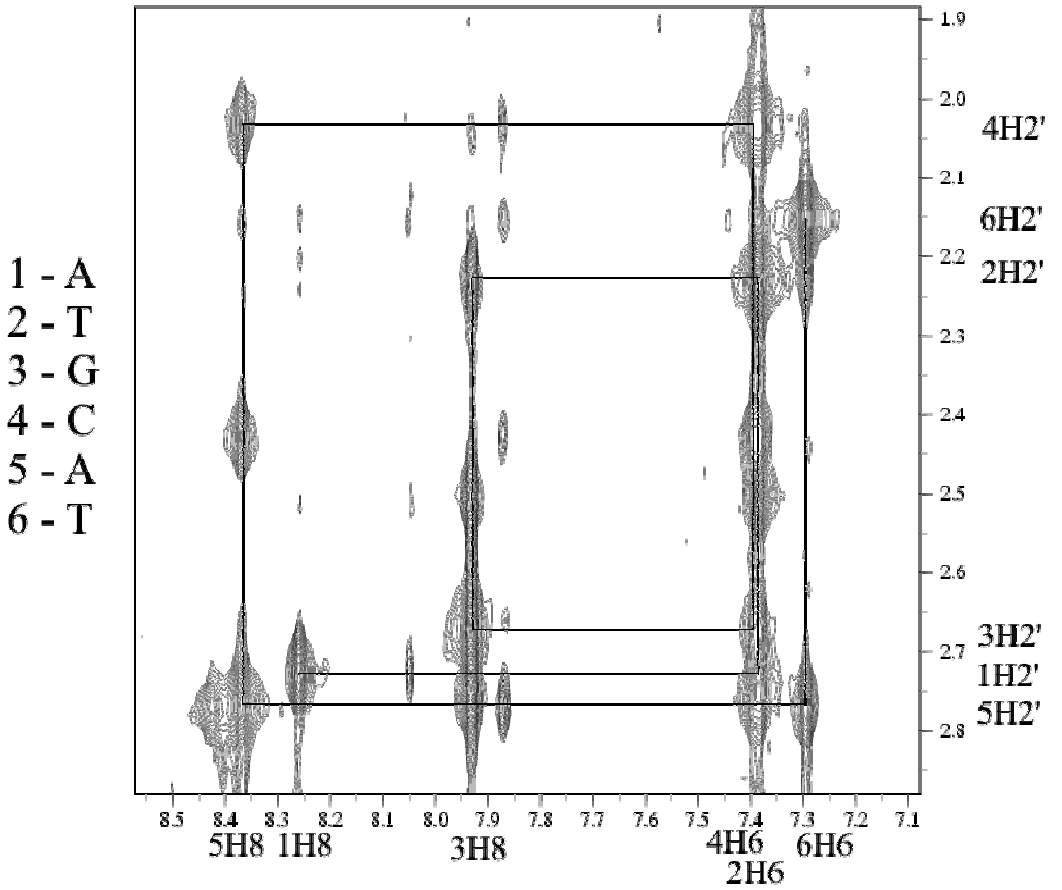

Fig. 3. *The sequential assignment procedure for H2'-H6/H8 resonances in ATGCAT hexamer.*

Materials and methods

Proton NMR spectra for duplex stability experiments were recorded at 4, 10, 15, 20 and 25 C on Varian Unity INOVA 600 MHz spectrometer equipped with a cold probe. Determination of structures of all hexamers was achieved by analysis of two-dimensional nuclear Overhauser effect spectroscopy (NOESY) spectra collected at 4 C with 300 ms mixing time on samples dissolved in 90% H_2O/ 10% D_2O. The pH for all samples was adjusted to 7.0, the concentration was 2.5 mmol/L. The water peak was suppressed by WATERGATE 3-9-19 pulse sequence [3]. To get good water suppression the carrier frequency was set on the water resonance. All spectra were acquired with 2048 and 512 complex points giving t_{2max}=0.164 s and t_{1max}=0,085 s in the F2 and F1 dimensions, respectively. The relaxation delay d1=1 s was used. All spectra were referenced to water signal (δ=5.00 ppm). Time domain data were zero filled twice in both dimension, multiplied with cosine window function and processed with PROSA [4]. Further analysis was done using XEASY [5]. Peak volumes from XEASY were converted into distance constraints and fed to CYANA [6] for geometry optimization using simulated annealing and torsion angle dynamics. CYANA output was directed to OPAL [7], a molecular energy minimization program. The output of statistical results is presented in Table 1. The energy-minimized molecular structures were analyzed using 3DNA software package [8].

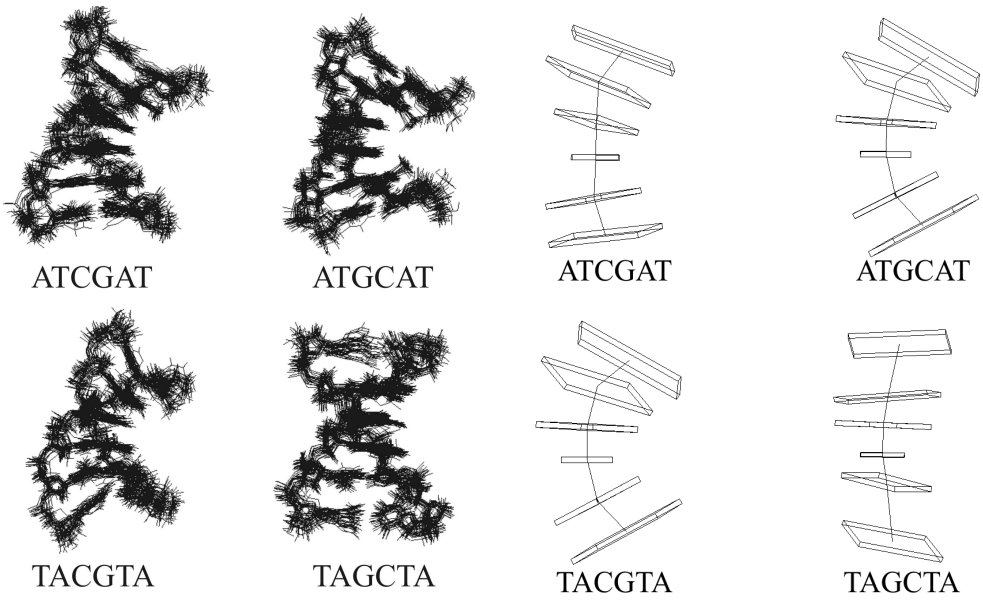

Fig. 4. *The superposition of 20 structures of each hexamer and schematic helix representations.*

Results and Discussion

Resonance assignment and interproton distance determination

To assign resonances we used exclusively ^{1}H-^{1}H NOESY spectroscopy. The NOE peak intensity is inversely proportional to the sixth power of the distance between the protons. So the nearest protons will give large, easy observable NOEs. For sequential assignment we used distances shown at Fig. 2. For example, the H2'-H6/H8 assignment pathway is shown in Fig. 3. Each H6/H8 proton gives cross-peak with H2'/H2" of its own residue and of previous

residue. Using this fact it is possible to go through all structure and assign all H6/H8 and H2'/H2" protons. The assignment of H1' and H3' protons were accomplished in the similar manner using the corresponding sequential connectivities separately [9]. The assignment of mobile NH protons gives information about the hydrogen bonds that connecting both strands of DNA duplex together. The assignment procedure and peak integration were performed by program XEASY. After the assignment and integration all peaks the data necessary for structure calculation were collected.

Three-dimensional solution structures calculation

Geometrical structures were calculated from peak intensities by CYANA, so that distance limit constraint violations are less than 0.1 Angstroms. CYANA generates upper and lower limit distant constraints from the peak integral file, generated by XEASY. The standard B-DNA angle constraints around phosphorous (alpha, beta, zeta and epsi angles) were introduced allowing for +/- 30 degree extra flexibility. Watson-Crick hydrogen bond constraints were used too. 20 best structures out of 100 calculated were used for energy minimization. To get final energy-minimized three-dimensional structures, the energy-minimization program OPAL was used. Files from CYANA were transformed into OPAL file format and minimized in water solution. Each structure was immersed in water box by generating ~ 400 water molecules in such way that thickness of water layer around the DNA was ~ 6 angstroms. Statistics from OPAL are shown in Table 1. The AMBER energies [10] are sufficiently low, there is almost no distance constraint violation, so, structures can be considered to be qualitative. Superpositions of all 20 structures of every hexamer are shown in Fig. 4. For structure analysis only central four base pairs were used, because of larger mobility of terminal bases.

Table 1.

Structural statistics from CYANA and OPAL.

Parameter	***ATCGAT***	***ATGCAT***	***TACGTA***	***TAGCTA***
Nonredundant NOE upper distance limits	132	178	131	218
Stereospecific assignments	5	5	6	4
Dihedral angle restraints	69	69	69	69
Physical AMBER-energies (kcal/mol) of				
solute-solute interactions	29.42 ±17.69	8.19 ± 10.11	-10.56 ± 9.77	18.63 ± 11.93
solute-water interactions	-2309.13 ± 71.09	-2246.20 ± 59.48	-2330.51 ± 33.77	-2256.92 ± 47.27
water-water interactions	-4414.98 ± 181.06	-4243.78 ± 157.61	-4730.81 ± 108.76	-4388.10 ± 109.80
all interactions	-6694.68 ± 216.07	-6481.79 ± 191.08	-7071.89 ± 115.09	-6626.40 ± 109.72
NOE constraint violations				
Number > 0.10 Angstroms	0.00 ± 0.00	1.95 ± 0.38	0.00 ± 0.00	1.50 ± 0.50
Sum (Angstroms)	1.71 ± 0.21	3.24 ± 0.15	2.43 ± 0.19	4.35 ± 0.17
Maximum (Angstroms)	0.09 ± 0.00	0.10 ± 0.00	0.09 ± 0.00	0.11 ± 0.00
Dihedral angle constraint violations				
Number > 0.25 degrees	0.35 ± 0.57	0.15 ± 0.36	0.00 ± 0.00	0.00 ± 0.00
Sum (degrees)	8.58 ± 1.89	6.86 ± 1.13	2.80 ± 2.07	5.14 ± 1.70
Maximum (degrees)	2.46 ± 0.12	2.35 ± 0.15	1.05 ± 0.55	1.52 ± 0.24
Rmsd (Angstroms)	0.93 ± 0.14	0.85 ± 0.13	1.05 ± 0.19	0.79 ± 0.14

Table 2.

Base pair and step parameters.

ATCGAT

Step	Rise	h-Twist	Prop-Tw	Roll	Incl.
T-A	3.57	29.07	-3.52	0.76	2.19
C-G	3.71	33.26	8.33	7.49	13.18
G-C	3.35	31	10.03	18.49	36.31
A-T	3.79	34.44	-4.74	8.57	15.01

ATGCAT

Step	Rise	h-Twist	Prop-Tw	Roll	Incl.
T-A	3.47	31.31	-13.14	1.12	2.08
G-C	3.83	40.35	-1.31	31.51	52.14
C-G	2.88	30.09	-1.86	3.99	7.92
A-T	3.66	36.97	-13.45	26.54	46.61

TACGTA

Step	Rise	h-Twist	Prop-Tw	Roll	Incl.
A-T	3.04	28.11	12.25	10.86	24.32
C-G	3.57	35.5	15.25	19.73	34.07
G-C	2.97	32.86	14.64	14.18	25.02
T-A	3.88	39.33	9.72	25.99	41.67

TAGCTA

Step	Rise	h-Twist	Prop-Tw	Roll	Incl.
A-T	5.14	41	-25.55	30.43	49.42
G-C	2.81	33.32	-3.5	-0.47	-0.82
C-G	3.04	39.48	-3.11	-1.9	-2.88
T-A	2.77	34.17	-28.33	1.84	3.04

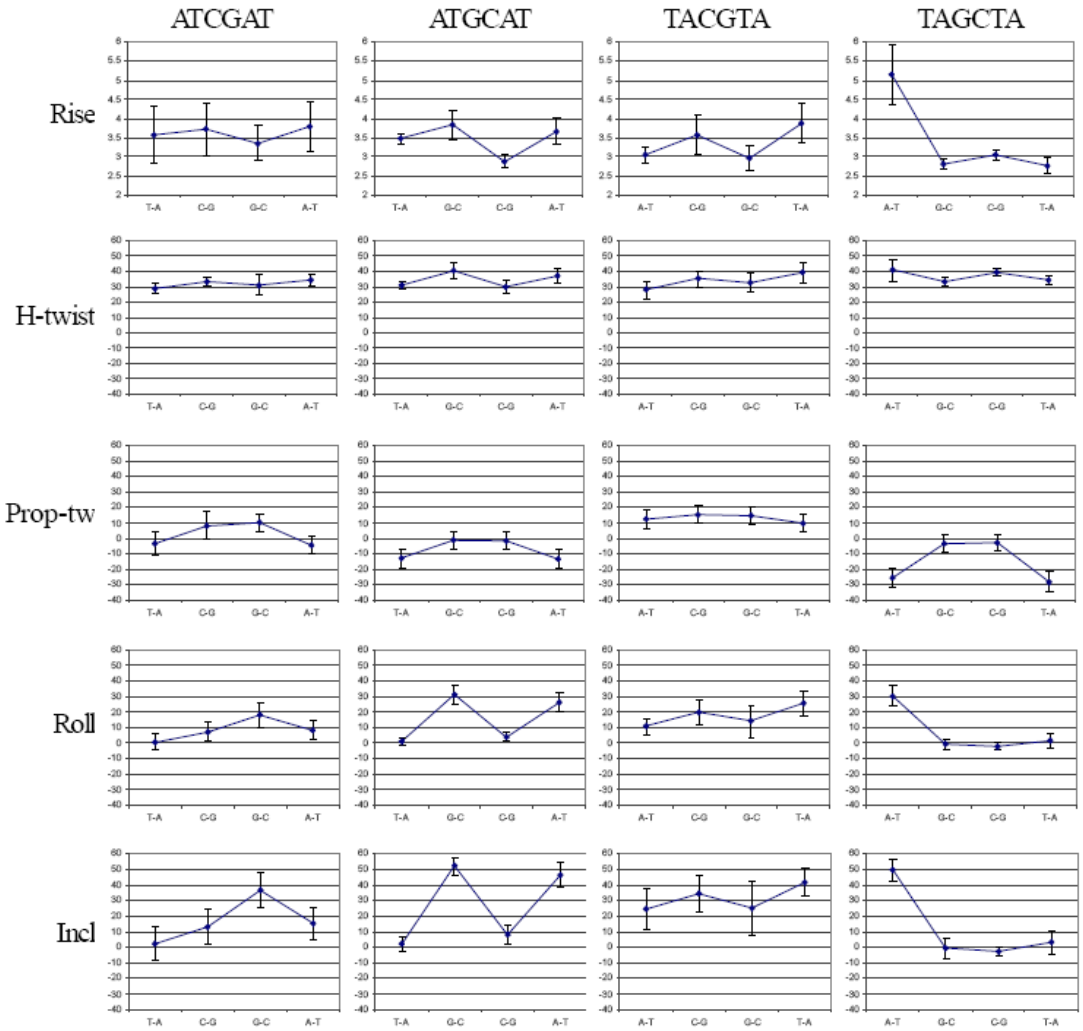

Fig. 5. *Graphical representation of central base step parameters.*

Helical and step parameters calculation

After the structure minimization, the base step and other helical parameters of all hexamers were determined. The parameter calculation procedure were performed for all 20 energy-minimized structures separately using 3DNA software package, afterwards the average values with standard deviations were calculated. Full list of step and helical parameters are shown in Table 2. The graphical representation of middle base steps parameters of every hexamer is shown in Fig. 5.

Schematic helix representations, showed in Fig. 4. were generated from the averaged helical parameters.

Structure analysis

The solution structure of all hexamers is that of right-handed B-DNA with the overall mean helical twist of ~ 36°. The macroscopic curvature, that naturally correlates with base step roll is largest for TACGTA (integral roll 70.8) and ATGCAT (integral roll 63.2) and lowest for TAGCTA (with integral roll 29.9). The large curvature in ATGCAT and low curvature in TAGCTA is consistent with the results obtained for the A-T/T-A steps and G-C middle base step in other analyzed oligomers structures [11].

Interestingly that the largest integral roll coinsides with the biologically active sequence ACGT in a central fragment that appears in the A, C, E, G boxes. These boxes are targets for sequence specific DNA-binding of oncoprotein c-Myc/Max heterodimer and that has been shown to directly transactivate the expression of a number of genes critical for cell proliferation, transformation, inhibition of differentiation and apoptosis. Probably such increased macroscopic curvature of ACGT sequence already in uncomplexed state allows for easier energy - saved interactions with proteins.

References

1. R.Foster, T. Izawa, N.-H. Chua. Plant bZIP proteins gather at ACGT elements, The FASEB J., **1994**, 8, 192-200.
2. A. J. Walhout, J. M. Gubbels, R. Bernards, P. C. van der Vliet, and H. T. Timmers. c-Myc/ Max heterodimers bind cooperatively to the E-box sequences located in the first intron of the rat ornithine decarboxylase (ODC) gene, Nucleic Acids Res. **1997**, 15, 1493-1501.
3. V.Sklenar, M.Piotto, R.Leppik, V.Saudek. Gradient-Tailored Water Suppression for $_{1}$H-$_{15}$N HSQC Experiments Optimized to Retain Full Sensitivity. *J. Magn. Reson.*, **1993**, *102A*, 241-245.
4. P. Güntert, V. Dötsch, G. Wider and K. Wüthrich. Processing of multi-dimensional NMR data with the new software PROSA, *J. Biomol. NMR*, **1992**, 2, 619-629.
5. Ch. Bartels, T.-H. Xia, M. Billeter, P. Güntert and K. Wüthrich. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules, *J. Biomolecular NMR,* **1995**, 5, 1-10.
6. P. Gьntert. Automated NMR protein structure calculation with CYANA. *Meth. Mol. Biol.,* **2004**, 278, 353-378.
7. P. Luginbьhl, P. Gьntert, M. Billeter, K. Wьthrich. The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. *J. Biomol. NMR,* **1996**, 8, 136-146.

8. L. Xiang-Jun and W. K. Olson. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, *Nucleic Acids Research*, **2003**, 31, 5108-5121.
9. G.M. Clore and A. M.Gronenborn. Sequence-dependent structural variations in two righthanded alternating pyrimidine-purine DNA oligomers in solution determined by nuclear Overhauser enhancement measurements, *The EMBO J.*, **1983**, 2109-2115.
10. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *JACS,* **1995***,* 117, 5179-5197.
11. D. S. Goodsell, M. Kaczor-Grzeskowiak and R. E. Dickerson. The Crystal Structure of CC-A-T-T-A-A-T-G-G. Implications for Bending of B-DNA at T-A Steps, *J. Mol. Biol*., **1994**, 79-96.
12. K. Wüthrich. *NMR of proteins and nucleic acids,* Wiley, 1986, 292 pp.
13. S. Neidle. *Nucleic acid structure and recognition,* Oxford University Press, 2002, 188 pp.

K. Zinovjevs, K. Jaudzems, E. Liepiņš. DNS heksamēru (ATGCAT, ATCGAT, TAGCTA, TACGTA) 3D struktūras noteikšana, izmantojot 2D kmr spektroskopiju *Noteiktas struktūras izšķīdinātiem paškomplementāriem DNS heksamēriem ar ApT vai TpA bāzes soļiem katrā galā un ar CpG vai GpC soļiem vidū, kuras atgādina bioloģiski svarīgās "E-box" sekvences. Visi četri heksamēri veido labēji virzītu antiparalēlu spirālveida dupleksa otrējo struktūru. 1H-KMR spektru temperatūras pētījumi parādīja, ka izpētītie dupleksi ir stabīli vismaz līdz +25°C. Ārējo bāzu pāru palielinātas mobilitātes dēļ DNS standartparametri aprēķināti un izanalizēti tikai katra dupleksa četriem vidējiem soļiem . Makroskopiskās izliektības analīze uzrādīja secību rindā TACGTA>ATGCAT>ATCGAT>TAGCTA. Secināts, ka palielināta makroskopiska izliektība jau brīvā nesaistītā stāvoklī varētu veicināt vieglāku, enerģētiski izdevīgāku DNS lokālo fragmentu mijiedarbību ar proteīniem vai ārstniecības preparātiem.*

K. Zinovjevs, K. Jaudzems, E. Liepins. Determination of the 3D structure of DNA hexamers (ATGCAT, ATCGAT, TAGCTA, TACGTA) BY 2D NMR spectroscopy. *The solution structures of self-complementary DNA hexamers with ApT or TpA base steps on each end and with CpG or GpC base steps in the center resembling the biologically important "E-box" sequences were determined. All four hexamers form the antiparallel-stranded right-handed duplexes. The 1H-NMR spectra acquired at higher temperatures (up to 25° C) showed that all duplexes are stable. The base-pair step parameters were calculated for four middle steps of all duplexes. The analysis of macroscopic curvature gave order TACGTA>ATGCAT>ATCGAT>TAGCTA. It was proposed that the increased macroscopic curvature of sequence already in uncomplexed state allows for easier energy – saved interactions with proteins or drugs.*

К. Зиновьев, К.Яудземс, Э.Лиепиньш. Определение пространственной структуры гексамеров ДНК (ATGCAT, ATCGAT, TAGCTA, TACGTA), с использованием двумерной спектроскопии ЯМР. *Определены структуры самокомплементарных гексамеров ДНК в растворе, с основаниями ApT или TpA на концах и CpG или GpC в центре, отвечающие биологически важной последовательности "Ebox". Все 4 гексамера образуют антипараллельный, правовинтовой дуплекс. 1H-ЯМР спектры, снятые при повышенной температуре (до 25° С) показали, что дуплексы стабильны. Шаговые параметры пар оснований были вычислены для 4 центральных пар каждого дуплекса. Анализ макроскопического искривления спирали показал порядок TACGTA>ATGCAT>ATCGAT>TAGCTA. Предположено, что макроскопическое искривление уже в свободном состоянии способствует более лёгкому, энергетически выгодному взаимодействию локального фрагмента ДНК с протеинами или лекарственными препаратами.*