RUSSIAN ACADEMY OF SCIENCES
SIBERIAN BRANCH

INSTITUTE OF CYTOLOGY AND GENETICS
LABORATORY OF THEORETICAL GENETICS

PROCEEDINGS OF THE SECOND INTERNATIONAL CONFERENCE ON BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE

Volume 3
INTAS section

BGRS'2000
Novosibirsk, Russia
August 7-11, 2000

ICG, Novosibirsk, 2000

International Program Committee

Nikolay Kolchanov, Institute of Cytology and Genetics, Novosibirsk, Russia (Chairman of the Conference)
G. Christian Overton, Center for Bioinformatics, University of Pennsylvania, USA (Co-Chairman of the Conference)
Ralf Hofestadt, University Magdeburg, Germany (Co-Chairman of the Conference)
Patrizio Arrigo, Institute of Electronic Circuits, CNR, Italy
Martin Bishop, Human Genome Mapping Project Resource Centre, UK
Philip Bourne, SDSC, San-Diego, USA
Philipp Bucher, Swiss Institute for Experimental Cancer Research, Switzerland
Chris Burge, MIT Center for Cancer Research, Cambridge, MA, USA
Julio Collado-Vides, National University of Mexico, Mexico
Jim Fickett, SmithKline Beecham Pharmaceuticals, USA
Mikhail Gelfand, Institute of Protein Research, RAS, Moscow, Russia
Charlie Hodgman, GlaxoWelcome Research Medicine Center, UK
Minoru Kanehisa, Kyoto University, Kyoto, Japan
Kotoko Nakata, National Institute of Health Sciences, Tokyo, Japan
Leonid Kalinichenko, Institute of Problems of Informatics RAN, Moscow, Russia
Lev Kisselev, Engelhardt Institute of Molecular Biology, Moscow, Russia
Luhua Lai, Institute of Physical Chemistry, Peking University, Beijing, China
Hwa A. Lim, D'Trends, Inc, USA
Gerhard Michal, Tutzingen, Germany
George Michaels, Genomics Lead Development, Monsanto Co.,USA
Luciano Milanesi, ITBA, Milan, Italy
Andrey Mironov, State Center for Applied Genetics, Moscow, Russia
Ken Nishikawa, Center for Information Biology, National Institute of Genetics, Japan
Manuel Peitsch, Glaxo Wellcome Experimental Research SA,Geneva, Switzerland
Mikhail Ponomarenko, Institute of Cytology and Genetics, Novosibirsk, Russia
Vadim Ratner, Institute of Cytology and Genetics, Novosibirsk, Russia
John Reinitz, Mt. Sinai Med. School, USA
Aida Romashchenko, Institute of Cytology and Genetics, Novosibirsk, Russia
Akinori Sarai, RIKEN Tsukuba Life Science Center, Tsukuba, Japan
Victor Solovyev, The Sanger Centre, Cambridge, UK
Masaru Tomita, Bioinformatics Laboratory of Keio University, Japan
Eduard Trifonov, Weizmann Institute of Science, Rehovor, Israel
Vladimir Tumanian, Engelhardt Institute of Molecular Biology, Moscow, Russia
Edgar Wingender, GBF, Braunschweig, Germany
Michael Zhang, Cold Spring Harbor Laboratory, Cold Spring Harbor, USA

Local Organizing Committee

Galina Kiseleva, Institute of Cytology and Genetics, Novosibirsk
Dmitry Afonnikov, Institute of Cytology and Genetics, Novosibirsk
Vasily Areschenko, Siberian Branch of RAS, Novosibirsk
Elena Borovskikh, Institute of Cytology and Genetics, Novosibirsk
Dmitry Grigorovich, Institute of Cytology and Genetics, Novosibirsk
Nadya Omelianchuk, Institute of Cytology and Genetics, Novosibirsk
Andrey Kharkevich, Institute of Cytology and Genetics, Novosibirsk
Anatoly Kushnir, Institute of Cytology and Genetics, Novosibirsk
Anatoly Kurbatov, Institute of Archeology and Ethnography, Novosibirsk
Sergey Lavryushev, Institute of Cytology and Genetics, Novosibirsk
Yuri Orlov, Institute of Cytology and Genetics, Novosibirsk
Galina Orlova, Institute of Cytology and Genetics, Novosibirsk

INTRODUCTION

Three volumes of Proceedings of the International Conference BGRS-2000 encounting about 180 abstracts are aimed to direct an attention to the actual problems in bioinformatics of genome regulation and structure. The Conference BGRS-2000 organized by the Laboratory of Theoretical Genetics of the Institute of Cytology and Genetics of Siberian Branch of Russian Academy of Sciences will be held in Novosibirsk, Russia, in August 7-11, 2000. This Conference will be the second in the series: the First International Conference on Bioinformatics of Genome Regulation and Structure – BGRS-98 was held in Novosibirsk in August 1998.

The question may arise: Why the Conferences BGRS attract their attention directly to the problems dealing with genome regulation and structure? The answer could be as follows: the structure and regulation of genome are the counterparts of life at molecular level; that is why understanding of fundamental principles of regulatory genomic machinery is impossible unless their structural organization is known, and *vice versa*.

During two years that have passed from the first BGRS Conference, the experimental genome study including applications to direct sequencing and mapping became of ever-growing scale. The huge bulk of experimental data on nucleotide sequences of complete bacterial genomes, the sequencing of which became a routine procedure in molecular biology, are being accumulated. Besides, complete genome of Drosophila is being deciphered, and human genome sequencing is drawing towards completion.

The ever-growing impact in genome studying is produced by novel experimental techniques. In particular, the EST technique is widely used in studying gene structure and gene expression patterns. Besides, microarray methods aimed at extracting unique and complete information on genome functioning and enabling to study simultaneously the expression patterns of dozen thousands of genes including those obtained at a single cell level, become implemented massively. In addition, single nucleotide polymorphism (SNP) technique provides a huge bulk of experimental data for studying regularities in mutation-assisted genome variability. Large-scale proteomic initiatives in the near future will lead to accumulation of large massifs of information on structure-functional organization of proteins.

The huge volume of experimental data that has been acquired on genome structure, functioning and gene expression regulation demonstrate the blistering growth. Development of informational-computational technologies of novel generation is a challenging problem of bioinformatics. Bioinformatics has entered that very phase of development, when decisions of the challenging problems determine the realization of large-scale experimental research projects directed to studying genome structure, function, and evolution.

By analyzing the papers submitted for publication in the three-volume issues of the BGRS-2000, the Organizing Committee came to a conclusion that participants of the Conference have concentrated their attention at consideration of the hottest items in bioinformatics listed below:

(1) Development of the novel generation of databases providing more complex, deep, and comprehensive description of (i) genome structure, function, and evolution, (ii) regulatory genome sequences, (iii) regulatory proteins, (iv) genetic networks, (v) signal transduction pathways and genetically controlled metabolic pathways.

(2) Development of computer technologies for automated knowledge discovery and data mining in the databases: ultra-rapid experimental methods developed for extracting molecular-biological data should correspond to similarly advanced technologies designed for automated treatment of these data, these technologies enabling to get the maximum of reliable and significant knowledge about genome function, regulation, and structure out of computer databases.

(3) Development of rigorous scientific methods for analysis of gene structure, discovery and modeling. Along with traditional approaches based on recognition of potential regulatory sites and coding regions, and their combinations, the more resolving power in solving this problem is demonstrated by the approaches based on comparative genomics, including both data search throughout databases and comparison of extended genome regions and even complete genomes (in case of bacteria).

(4) Development and improvement of methods in comparative genomics that became one of the most high-powered and perspective directions in modern bioinformatics. The efficient algorithms developed within the frames of comparative genomics appear to be more reliable tool acquired for gene recognition and gene reconstruction throughout *de novo* sequenced genome DNA, for recognition of regulatory elements controlling genome functions and gene expression regulation. Besides, comparative genomics will go a long way towards revealing fundamental principles of genome organization and regularities in genome molecular evolution.

(5) Further mastering of approaches designed within the frames of comparative genomics strongly depends upon comprehension of fundamental regularities in genome organization and evolution. That is why computer analysis and modeling of genome mutability, together with studying of fundamental laws of evolution of genomes, coding gene regions and regulatory genomic sequences become the matter of especial importance. Accumulation of knowledge in this field will certainly help in searching for objective methods in annotating and finding of genes and regulatory signals in genomic sequences.

(6) Development of novel generation of mathematical algorithms implemented for analysis of regulatory genome sequences (RGS) and for accounting of real complexity of RGS. These algorithms are characterized by a large variety of parameters significant for gene functioning, by blockwise structure, and hierarchy in RGS organization. On the background of these algorithms, the fine accuracy methods are being developed for

recognition and prediction of quantitative values of regulatory genomic sequences activity of various types, which provide implementation of numerous genome functions regulating basic stages of gene expression.

(7) Revealing of fundamental regularities in structure-functional organization of RGS controlling basic types of molecular-genetical processes (i.e., replication, transcription, splicing, polyadenylation/processing, translation, etc). Besides revealing the regularities in structure-functional organization of RGS that are valuable for increasing the accuracy of their recognition, this analysis allows to obtain a fundamental knowledge on molecular mechanisms of RGS functioning, thus enabling to solve one of the main problems in bioinformatics of genome regulation and structure.

(8) Development of methods aimed at prediction and recognition of structure-functional organization of proteins encoded by the genes detected within *de novo* sequenced genome sequences. The lack of unified technological production line processing from the coding gene regions in the sequenced genomes to prediction of structure-functional organization of proteins encoded by these genes serves as the stopping brakes for implementation of large-scale genome projects. It should be stressed that during the solving of the task, a large attention should be paid to detecting fundamental principles of protein organization and evolution. The most important is the studying of aspects of protein function and structure related to genome regulation. During the recent years, the tendency manifested itself in convergence and intersection of the lines in bioinformatics of genome regulation and structure and in computer-assisted proteomics. This observation is clearly approved in Proceedings of BGRS-2000.

(9) Large-scale genome analysis. The other day computer analysis was restricted to studying of local context regularities in genome structure. Currently, due to widespread sequencing of complete genomes and their extra-extended fragments, a possibility first appeared to analyze large-scale context dependencies in genome DNA organization. To tackle this problem, it is necessary to develop operative methods aimed at analysis of extra-extended genome sequences.

(10) Description in databases and modeling of genetical networks, which control the processes of basic metabolism, cell division and differentiation, organ- and tissue morphogenesis, growth and development of an organism; support of homeostasis of molecular, biochemical, and physiological parameters of organisms, etc. Systemic investigation of mechanisms related to genome functioning and gene expression regulation at the level of gene networks and signal transduction pathways should be provided. On the grounds of these very processes, the key problem in bioinformatics, that is, recognition of phenotypical characteristics of an organism on the basis of information encoded in their genomes will be solved in future.

(11) Development of efficient technologies for integration of informational and software resources on the structure and regulation of genomes and designing on this basis of super-large computer systems implemented for analysis and modeling of intricate molecular-genetic systems and processes.

(12) Analysis of fundamental regularities in (i) genome functioning, organization, and evolution, (ii) the mechanisms governing the coding of genetical information, (iii) molecular bases of realization of genetical language, principles of organization, functioning, and evolution of genetical networks and molecular-genetic systems.

All the questions listed above will be suggested to consideration of participants of BGRS'2000 at 7 sections and presented in a form of plenary lectures, oral communications, posters, Internet computer demonstrations and round table discussions.

BGRS'2000 will bring together the experts in Bioinformatics to discuss the progress in the field of bioinformatics of genome regulation and structure achieved at the end of 20th century, the basic approaches devoted (i) to data description and analysis; (ii) modeling of complex molecular-genetical systems; (iii) to revealing of fundamental principles of genome organization and evolution and of mechanisms of genetical information coding; (iv) to evaluation and marking off the future trends in this field.

The researchers working in the fields of experimental biology and interested in application of Bioinformatics methods in their work are also the participants of the Conference. With this respect, the Conference is expected to be a stimulating event not only giving a future development of bioinformatics as it is, but also establishing new links between Bioinformatics and experimental research.

By working out the BGRS2000 schedule, the Organizing Committee has tried to keep the balance between technical (applied) and fundamental aspects in bioinformatics. This principle has a clear reflection in contents of Proceedings of the Conference. Herein, we have tried to follow the well-known and far-back principle: «nothing is more practical than the good theory».

Professor Ralf Hofestadt
Co-Chairman of the Conference
University Magdeburg, Germany

Professor Nikolay Kolchanov
Co-Chairman of the Conference
Vice-Director of the Institute of Cytology and Genetics
Novosibirsk Russia

G. Christian Overton

15.02.1948 – 1.06.2000

The death of Dr. Overton, premature and unexpected, is a body blow for all Bioinformatics community and for us, his friends and colleagues in Russia. Chris always paid a great attention to strengthening of the international cooperation in the Bioinformatics research. That is why, he left behind along with brilliant scientific results an example of real international collaboration in science. As an example of international cooperation initiated by Chris may serve the collaboration with Laboratory of Theoretical Genetics of the Institute of Cytology and Genetics of Siberian Branch of Russian Academy of Sciences. From 1995, this laboratory and the Center for Bioinformatics at University of Pennsylvania together made a common research despite of distance between Philadelphia and Novosibirsk, and we are very grateful to Chris for his support, assistance and understanding friendly provided.

Dr. G. Christian Overton was the founding Director of the Center for Bioinformatics at Penn, established in 1997 as an interdisciplinary venture between the Schools of Medicine, Arts and Sciences, and Engineering and Applied Science. He was also an Associate Professor in the Department of Genetics, and held a secondary appointment in the Department of Computer and Information Science in the School of Engineering and Applied Science. Dr. Overton received his Bachelor of Science degree in Mathematics and Physics from the University of New Mexico in 1971, his Ph.D. in Biophysics from the Johns Hopkins University in 1978, and his M.S.E. in Computer and Information Science from the University of Pennsylvania in 1986. After receiving his M.S.E, he returned to the University of Pennsylvania in 1991 as an Associate Professor. In addition to his research, Dr. Overton was an Editor for the Journal of Computational Biology, Bioinformatics, and Gene/Gene-COMBIS as well as the Member of the Board of Directors for the International Society for Computational Biology.

Dr. Overton's brilliant skills in biophysics and bioinformatics, his deep understanding of the challenges in biology, medicine, and computer science enabled him to organize many outstanding research projects, which bridge the gap between experimental biology and computer science aimed to experimental data treatment. Dr. Overton is internationally recognized as a pioneer in genomic research and application of computational approaches for solving biological problems. He focused on problems associated with database integration, genome annotation, gene recognition, and detection of regulatory elements governing the expression of many genes that comprise the human genome.

Chris was one of the Co-Organizers of the BGRS2000 Conference. Due to his activity, the Organizing Committee managed to put together the effort of those interested in the basic approaches and trends in bioinformatics. Chris will be remembered for his love of science, his charm, good nature and his great intelligence. For people here in Novosibirsk who met him and knew him well he will be remembered as a faithful and good friend and as a researcher devoted to science.

Professor Nikolay Kolchanov,
Head of Laboratory of Theoretical Genetics,
Vice-Director of the Institute of Cytology and Genetics,
Novosibirsk, Russia

CONTENTS

BGRS 2000

INTAS SECTION

PREFACE

Over the recent years genome research using direct sequencing and mapping has been advancing at ever increasing rates. The primary structure of thousands of genes, and numerous fragments of genomic DNA with unknown function have been brought to the researcher's attention by mass sequencing. The huge volume of experimental data that has been acquired on regulation of genome function and the mechanisms of gene expression regulation continues to grow. The bulk of experimental data on splicing, polyadenylation/processing, and translation is also impressive. Systemic investigation of the mechanisms of genome function and gene expression regulation at the level of gene networks and signal transduction pathways attracts more and more attention of researchers.

That is why the Organizing Committee has paid a special attention to invitation of researchers working in the fields of experimental biology and interested in application of Bioinformatics methods to participate in BGRS'2000. Due to these purposes, the Organizing Commettee decided to organize a special INTAS Session within the frames of the Conference. At this Session, the works of Russian scientists in the fields of experimental genomics, molecular biology and genetics, along with the studies in computer genomics supported by INTAS grants will be presented.

We expect that the INTAS Session will be a stimulating event to establish new links between Bioinformatics and experimental research, believe that thhis event will promote further progress in Bioinformatics and Biology, and will be interesting for both bioinformatics community, and those many biologists who investigate the structural and functional organisation of pro- and eukaryotic genomes by experimental methods.

The INTAS promotes international scientific co-operation and has already funded over 350 projects in the different fields biology. The Organizing Committee is very grateful to INTAS for the decision to contribute to the costs of BGRS'2000. INTAS monitoring day is integrated into the overall program of the conference. This day includes the session with oral presentations of INTAS progress reports. The poster session, computer demonstrations on INTAS projects or related to them objectives and round table discussions on possible application of Bioinformatics approaches to realization of both INTAS and other projects are also in the Program of the conference. The special attention during discussions will be paid to the initiation of international concortia in these fields including collaboration via Internet by the development of data and knowledge bases and other computer approaches of the common interest on the base of contacts initiated during this conference. This special chapter contains the abstracts of presentations on monitoring of INTAS projects.

Professor Nikolay Kolchanov,
Head of Laboratory of Theoretical Genetics,
Vice-Director of the Institute of Cytology and Genetics,
Novosibirsk, Russia

LIFE SCIENCES AT INTAS

The international science funding organisation INTAS promotes scientific co-operation between scientists in its member states and the New Independent States of the former Soviet Union. Its membership comprises the European Union member states, Iceland, Israel, Latvia, Norway, Romania, Slovenia, Switzerland and the European Community.

INTAS has already funded almost 2,150 research projects in a wide variety of scientific disciplines. Over 350 of these projects relate to microbiology, physiology, biochemistry & physics, genetics, medicine, biotechnology… Furthermore, some 16 conferences and 47 NIS young scientists working in the field of life science have already enjoyed INTAS funding. Could you be next?

INTAS Calls for proposals

There is at least 30 MEuro available for joint research projects and networks in 2000 & 2001 through two Open Calls, a Joint INTAS-CERN Call and Thematic Calls. The following will probably interest researchers in life sciences most:

- **Food Call 2000 -** 1.5 MEuro for joint research relating to the whole food chain, from production to consumption.
- **Information Call 2000 -** 1.5 MEuro for joint research relating to the storage, processing, retrieval, dissemination… of electronic information in all scientific fields.
- **Open Call 2001 -** 13 MEuro for joint research and networks in Life and Earth Sciences and Physics (Condensed Matter, Plasma, Optics…)

The Calls 2000 remain open till 29 September 2000 and the Call 2001 till 28 September 2001. For more information about the elibility criteria and the submission procedure, please check out the Information Package, available on http://www.intas.be or from INTAS by written request: infopack2000@intas.be;

fax: +32 2 549 0156 or Avenue des Arts 58, B-1000 Brussels, Belgium.

INTAS Young Scientist Programme

INTAS is eager to encourage NIS researchers under 35 in their scientific career. It therefore awards conference grants to young scientists involved in INTAS projects and has an extensive fellowship programme. There are project-linked, PhD, newly qualified post doctoral and experienced post doctoral fellowships.

1 MEuro is allocated to the Young Scientist Programme. There are 4 deadlines for the conference grants: 15 January, 15 April, 15 July, 15 October. There are 2 deadlines for the fellowships: 31 August 2000 and 31 March 2001.

For more details of this programme and application forms please consult the website or write to us by email (intas@intas.be), fax or regular mail.

Other actions

Other funding opportunities include grants to support international conferences involving 5 or more INTAS projects and infrastructure actions, which aim to support basic infrastructure which benefits a wider section of the scientific community, e.g. supplying technical literature, repairing research vessels on Lake Baikal…

For more information about these please consult the website or write to us at the above address.

Achievements in INTAS life science projects

As this scientific field covers a wealth of disciplines, the following projects are only a selection of some of the projects INTAS is funding but they reflect the wealth of disciplines:

- Correlates of extinction risk for Central Asian biodiversity;
- Structural organisation of Golgi compartment in Microspodians: one more example of a minimal secretory system?
- Mitochondrial dysfunction in neural disorders in the mammalian CNS;
- Methods, algorithms and software for functional and structural annotation of complete genomes;
- Advanced dental implants manufactured by selective laser sintering.

The life sciences are also responsible for many of the patents to result from INTAS projects. In 1999, in particular, half of the 24 patents were found in this scientific field. A number of the patents to date relate to the following:

- Replication of the genome of positive-strand RNA viruses;
- Maintenance & development of endangered culture collections of extremophilic microorganisms;
- Microbial preparations for plant protection and control of insect pests and diseases of agricultural crops.

INTAS' user-friendly tools

INTAS has done all it can to develop user friendly procedures and tools. In addition to the Internet submission and evaluation systems for the Calls for proposals, INTAS posts a summary of each project in its Catalogue of Projects and allows scientific institutes to link their homepage to the INTAS website. For more details please email: intas@intas.be.

INTAS has also opened an online partner search facility. To find out who is seeking which partners or to add your own request for partners, surf to Beyond INTAS on http://www.intas.be.

Ingmar Dirk Schmidt
INTAS, Scientific Officer, Belgium
e-mail: schmidt@intas.be

REGULATION OF MDR1 GENE EXPRESSION BY ds-OLIGONUCLEOTIDES

Shatskaya N.V., Bozhenok L.N., Chernolovskaya E.L., Vlassov V.V.

Novosibirsk Institute of Bioorganic Chemistry SB RAS, Novosibirsk, Russia
e-mail: witty5@mail.ru
*Corresponding author

Keywords: multiple drug resistance, oligonucleotides, transcription regulation

Resume

Development of cross-resistance to many anticancer drugs, termed multidrug resistance (MDR), is one of the major reasons why anticancer chemoherapy ultimately fails. This type of MDR is often associated with over-expression of MDR1 gene product, P-glycoprotein, a multifunctional drug transporter.

Control of MDR1 gene induction, perhaps can be achieved by using double stranded oligonucleotides (ds-ODN) mimicking transcription factors binding sites on the promoter region. These ODNs can inhibit gene activation via competitive interaction with the transcription factors. We have examined interaction of ds-ODNs mimicking Y-box site, activator binding site, GC-rich sequence and NF-IL6 binding site with the transcription factors during MDR1 induction.

We have found that A (A1) type complex forms on all four regions of promoter after activation of mdr1 expression and the complex formation does not depend on the nature of inductor while a number of minor complexes were drug-, region- and cell specific.

We tested these ODNs as potential inhibitors of MDR1 expression *in vivo*. Complete inhibition of MDR1 induction was achieved using ds-ODN mimicking Y-box or site of activator binding on Molt 3 cell line induced by doxorubicine. On HeLa cell line treated by vinblastine total inhibition of MDR1 expression was achieved by ODN mimicking the Y-box.

Introduction

Multidrug resistanse (MDR) is one of the main obstacles in the chemotherapeutic treatment of cancer. The product of the mdr1 gene (P-glicoprotein) acts as an energy-dependent efflux pump responsible for removal of a wide range of xenobiotics, including chemotherapeutic drugs.

One of perspective approaches for investigation of regulation of gene expression consists in using of double stranded oligodeoxynucleotides (ds-ODN) mimicking of promoter regions of genes competitive inhibiting the involved protein factors. Clusel at all (1993) have demonstrated that ds-ODN consisting a HNF-1 binding site can decrease the gene expression even at nanomolar concentration of ODN by forming complexes with transcription factors [1].

Methods

HeLa (human cervical carcinoma cell line), Molt3 (acute limfoblastoid leukemia) and K-562 (chronic myeloid leukemia), KB (epidermic carcinoma) were grown in IMDM with 10% embryonic serum in 5% CO2, 370C.

Drugs were added to cultural medium to concentration 5 nM- vinblastine, 10nM- doxorubicine, 60nM-cytarabine.

The transcriptionally active nuclear extract was prepared according to Dignam [2] with modification. The protein concentrations were measured according to Lowry [3].

Oligodeoxyribonucleotides were synthesized by standard phosphamide method on an automatic synthesizer and 5'-labeled with [[γ]-P32]ATP and polynucleotide kinase. Investigation of proteins binding to ds-ODN was made in buffer A: 4 mM glycerin, 1 mM MgCl2, 0.5 mM Na2EDTA, 0.5 mM DTT, 50 mM NaCl, 10 mM Tris-HCl, pH 7.5. Reaction mix (20mg of nuclear extract and 0.1-1×10-3 pmol ODN) was incubated for 10 min. at room temperature, then 15 min. in ice then 1/10 volume of sample 10- fold buffer (250 mM Tris-HCl, pH 7.5, 40% glycerin, 0.2% bromfenol blue) were added. Complexes were analyzed by electrophoresis.

Total cellular RNA was isolated by guanidine isothiocyanate method. Quantity of mRNA was measured by RT-PCR, as described by Murphy [4], β2- microglobuline was used as internal standard. The samples were analyzed by electrophoresis in 1xTris-borate buffer in10% PAAG.

Computer analysis of potential binding sites was made using program TESS (Transfac version 3.2), URL: http:/agave.humgen.uppen.edu/utess/tess32.

Results and discussion

We investigated in vitro interaction of transcription factors with four ds-ODNs mimicking regulators promoter regions of MDR1 gene involved in the process of the gene activation. The regions were chosen according to published data.

Sequence of promoter region and 20-mer oligonucleotides: NF- mimicking binding site of NF-IL6 factor (–151/-132), Y-b- mimicking Y-box site (-83/-64), GC- mimicking GC-rich region (-70/-49) and Act- mimicking activator binding site (-61/-42) are shown at Figure 1.

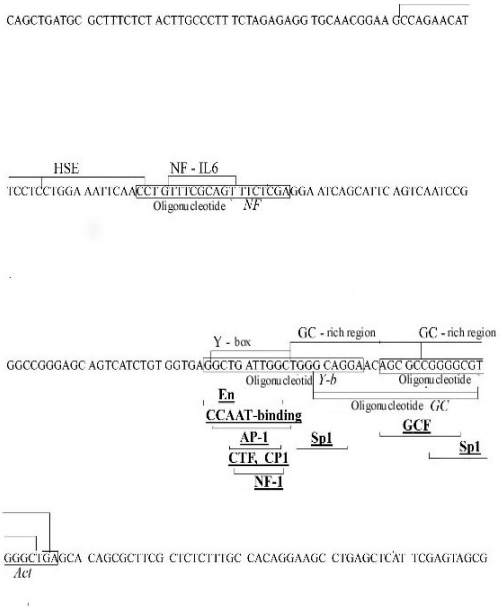


Figure 1. MDR1 promoter region

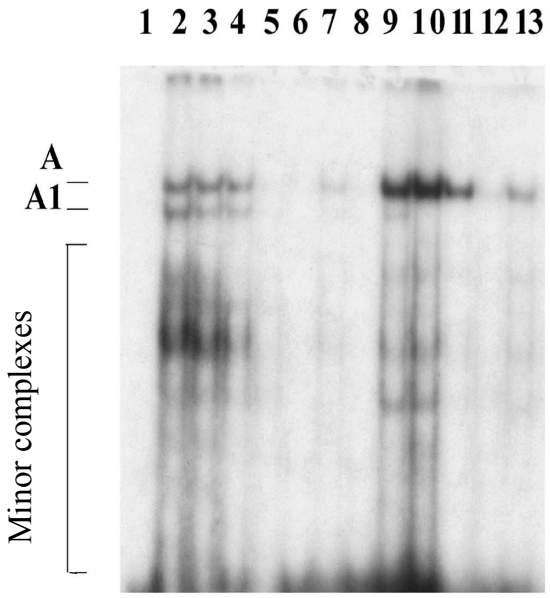


Figure 2. Gel-shift analysis binding oligonucleotide NF with proteins. Lane 1-control, 2-8 – extract from K-562 cell, 9-13 – extract from KB cell, 2, 9 – control, 3, 10, 4, 11, 5, 12 – doxorubicine induced cell, 6, 7, 8, 13 – cell cultivate with doxorubicine

Gel-shift analysis detects similar sets of protein complexes formed on the investigated promoter regions after short term drug induction. They can be divided into two types: the major complex, type A, and a number of the minor complexes (formation of A complex is typical for K-562 cells but it reduced after drug treatment).

It was found that A (A1) type complex is formed on all four regions of promoter and the complex formation does not depend on the nature of inductor. Minor complexes were drug-, region- and sell specific (Fig. 3A, 3B, lane 2, 5, 9, 12).

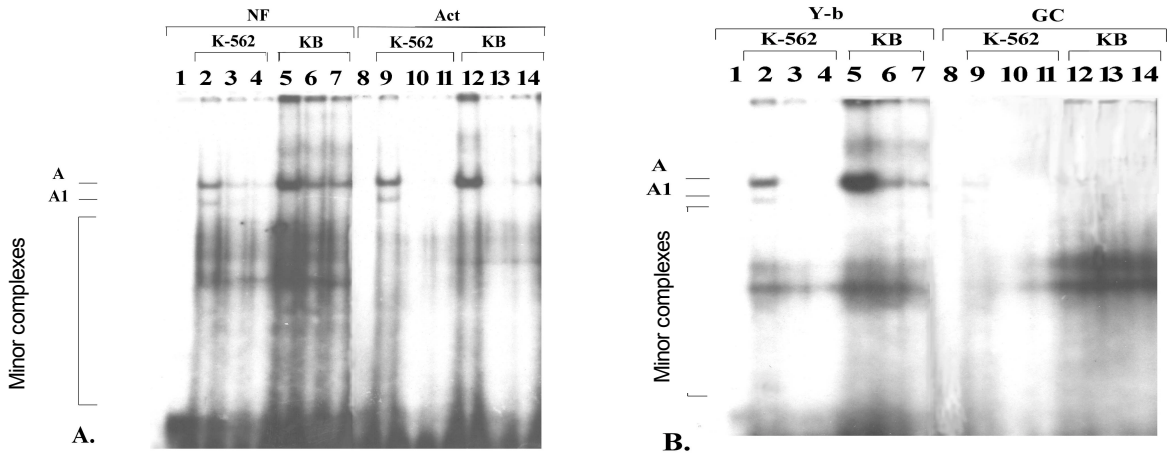


Figure 3. Gel-shift analysis of complex formation in extracts of K-562 and KB cells induced by cytarabine. Lane 1, 8 – oligonucleotides in complex formation buffer, 3, 6 – in the presence of 5 –fold excess oligonucleotide Act; 4A, 7A, 11- in the presence of 5-fold excess of oligonucleotide Y-b; 4B, 7B, 10, 13 – in presence oligonucleotide NF

Comparison of complex formation in the same region after gene activation by different drugs demonstrated that the most effective induction of the complex formation occur in the case of vinblastin treatment (Fig. 4, lane 6, 7, 13, 14).

These data correlate with higher RNA expression level in the vinblastine treated cells(Fig. 5).

In order to investigate possible interactions between different sequences of the regulating gene region in assembling of the main complex type $A(A1)_2$ we added to nuclear extract 5 fold excess of ds-ODN mimicking remoted sequence before incubating with [P^{32}] labeled ODN. It was found to inhibit the assembling of the main complex. Formation of some minor complexes was not effected (Fig. 3A, 3B, lane 3, 4, 6, 7, 10, 11, 13, 14). These data indicate that all four regions are involved in formation of complex A(A1).

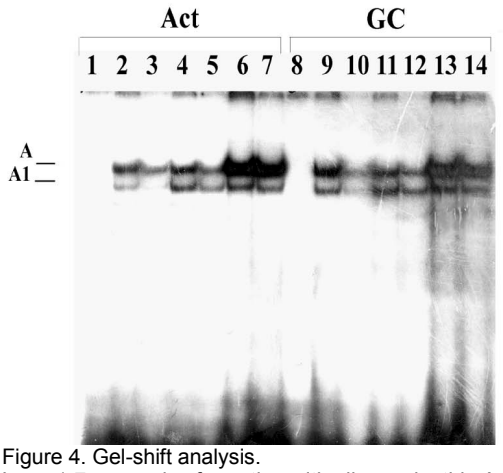


Figure 4. Gel-shift analysis.
Lane 1-7 – complex formation with oligonucleotide Act,
8-14 – with oligonucleotide GC,
1, 8 – oligonucleotide in complex formation buffer,
2, 3, 9, 10 – cell extract from cytarabine induced cell,
4, 5, 11, 12- cell extract from doxorubicine induced cell,
6, 7, 13, 14 – cell extract from vinblastin induced cell,
3, 5, 7, 10, 12, 14 – complex formation in presence of 10 excess of free oligonucleotide $(AT)_{16}$.

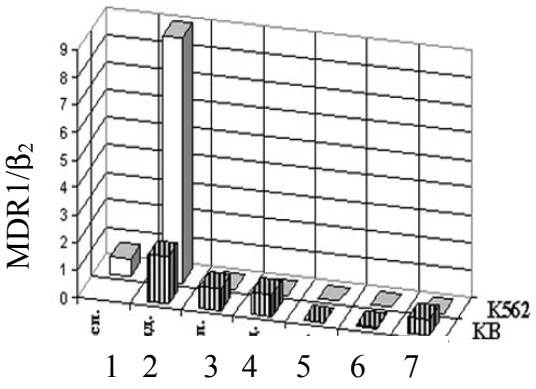


Figure 5.
1- resistant cell line grown in the presence of vinblastine
2- vinblastine treated cell line
3- resistant cell line grown in the presence of doxorubisine
4- doxorubicine treated cell line
5- resistant cell line grown in the presence of cytarabine
6- cytarabine treated cell line
7- control cells

The published data about cooperation of transcription factors or their subunits favour the second.

We have developed drug resistant cell sublines by growing cells in the media with increased concentrations of drugs. When we compared the binding of transcription factors with the regulatory regions in the wild type lines, induced by exposition to cytostatics and in drug resistant sublines (Fig. 2, lane 3, 6, 8, 10, 13), we have found that the most effective binding occurs in the first case. Quantity of the mdr1 mRNA decreases during prolonged drug treated (Fig. 5). Transferring of resistant cells to the medium with higher drug concentration result in increasing of the binding of protein to ds-ODNs (Fig. 2, lane 6, 7, 8). These data suggest that the trigger to MDR1 activation is the increase of the intracellular inductor concentration.

We have tested these ds-ODNs as specific inhibitors of gene expression *in vivo*.

Data are shown in Fig. 6.

Was have observed complete inhibition of MDR1 induction on Molt 3 cell line in the presence of doxorubicine by ds-ODNs mimicking Y-box or site of activator binding. ds-ODN mimicking Y-box caused increase of gene expression on control cells. In HeLa cell line treated by vinblastine total inhibition of the induction was achieved by ds-ODN mimicking Y-box. All four ds-ODNs had no substantial affects on gene expression on K-562 cell line.

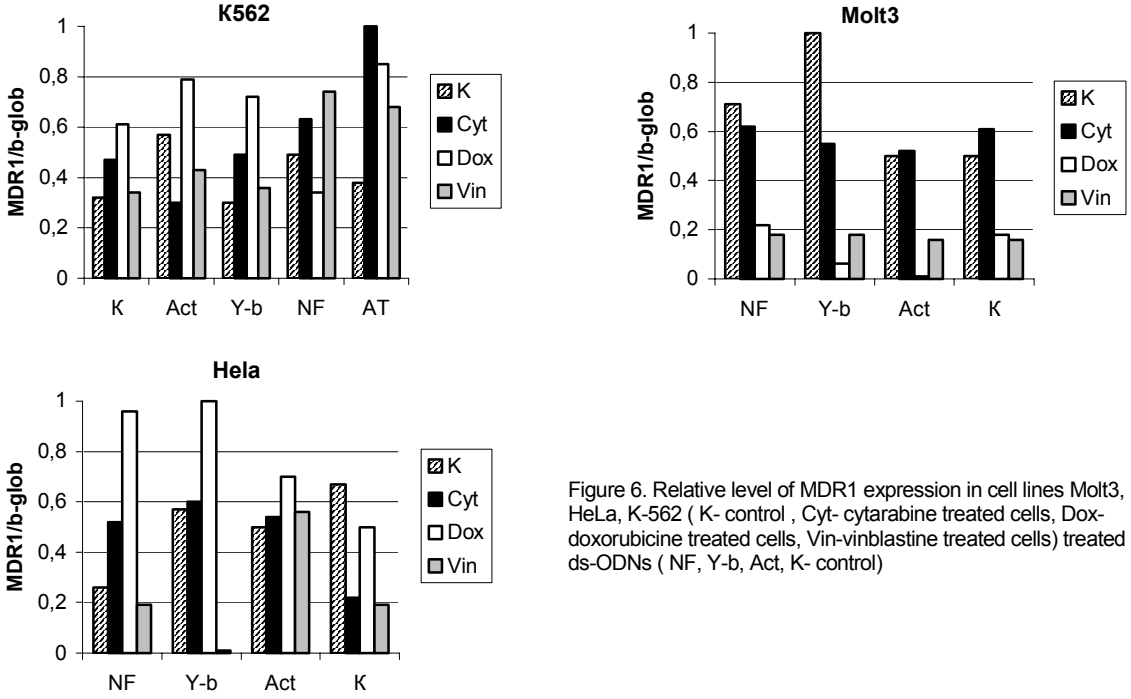


Figure 6. Relative level of MDR1 expression in cell lines Molt3, HeLa, K-562 (K- control , Cyt- cytarabine treated cells, Dox- doxorubicine treated cells, Vin-vinblastine treated cells) treated by ds-ODNs (NF, Y-b, Act, K- control)

The study is supported by the grant INTAS No. 99-630.

References

1. Clusel C., Ugarte E, Enjolras N., Vasseur M., Blumenfeld M. Ex vivo regulation of specific gene expression by nanomolar concentration of double-stranded dumbbell oligonucleotides // Nucleic Acids Res. 1993. V.21. P.3405-3411.
2. Dignam J. D., Lebovitz R. M., Roeder R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. // Nucleic Acids Res. 1983. Vol. 11. N. 5. P. 1475- 1483.
3. Lowry O. H., Rosenbrough N. J., Farr A. L., Randall R. J. Protein measurement with Foline phenol reagent // J. Biol. Chem. 1951. Vol. 193. P. 265 - 275.
4. Murphy L.D., Herzog C.E., Rudick J.B., Fojo A.J., Bates S.E. Use of the polymerase chain reaction in the quantitation of mdr-1 gene expression.// Biochemistry. 1990. V.29. P. 10351-55.

IDENTIFICATION OF OPTIMAL TARGET SITES FOR ANTISENSE OLIGONUCLEOTIDES IN THE α-SARCIN LOOP REGION OF *E. Coli* 23S rRNA

Petyuk V.A., Zenkova M.A., Vlassov V.V.

Novosibirsk Institute of Bioorganic Chemistry SB RAS, Novosibirsk, Russia
e-mail: petyuk@niboch.nsc.ru
*Corresponding author

Keywords: antisense oligonucleotides, RNA structure, α-sarcin loop

Resume

Motivation:

23S RNA of bacterial ribosomes represents an important potential therapeutic target. An apparent approach to targeting ribosomal RNA consists in using antisense oligonucleotides. Currently available theoretical approaches are of limited value for identification of optimal target sites for antisense oligonucleotides. Therefore systematic experimental studies are needed for unambiguous identification of the target sites.

Results:

Hybridization of antisense 15-mer oligonucleotides, complementary to the 2638-2682 region of *E. coli* 23S rRNA was investigated, using gel-mobility shift assay. 4 of the 7 tested oligonucleotides were shown to efficiently hybridize to the target RNA. Currently used theoretical approaches failed to predict the observed oligonucleotide binding affinities. Some correlation of the experimental results was observed with theoretically calculated affinities, when rearrangement of the RNA structure caused by oligonucleotide hybridization was taken into account.

Introduction

Natural RNAs usually are tightly folded and do not contain long single-stranded regions considered as ideal antisense oligonucleotide targets [Branch, 1998]. Binding of oligonucleotides to natural RNAs should be accompanied by rearrangements of RNA's structure. α-sarcin loop of 23S-type rRNA is the longest universally conserved sequence (12 nt) of rRNAs. The integrity of α-sarcin loop is absolutely important for ribosome function [Endo, 1982]. The structure of this site of 23S rRNA has been investigated in detail by methods of chemical and enzymatic probing, and by NMR [Szewczak, 1993] and x-ray crystallography [Correll, 1999]. We investigated hybridization of pentadecanucleotides with the fragment of *E. coli* 23S rRNA containing α-sarcin loop region to identify optimal oligonucleotide for targeting 23S rRNA.

Methods

171-nucleotide fragment of 23S rRNA (ECAS171) was prepared by *in vitro* transcription. Two synthetic DNA oligonucleotides (5'-TAATACGACTCACTATAGGCGCTGGAGAACTGAGGG-3' and 5'-GGGTCAGGGAGAACTCATCTCG-3') were used to amplify by RT-PCR the 2625-2795 fragment of 23S rRNA of *Escherichia coli* and to insert T7 promoter at the 5'-end for *in vitro* transcription.

5'-end ^{32}P labeling of ECAS171 RNA was performed with [γ–^{32}P]ATP and T4 polynucleotide kinase. Prior to 5'-end labeling, ECAS171 RNA was dephosphorylated using bacterial alkaline phosphatase. Uniformly labeled ECAS171 RNA was prepared by *in vitro* transcription using [α–^{32}P]ATP and T7 RNA polymerase.

Oligonucleotide hybridization was studied using gel mobility-shift assay [Petyuk, 1999]. Hybridization was performed in 50 mM Cacodylate-HCl pH 7.3, 0.1 mM EDTA, 300 mM KCl at 37°C for 1h. Positions of oligonucleotide binding sites were identified by RNase H cleavage assay.

Results

The target sites for the oligonucleotides complementary to the fragment of 23S rRNA, containing α-sarcin loop, are shown in Fig. 1. These sites represent an equidistant run with 5-nucleotide step, i.e. any site can be obtained by shifting the previous one by 5 nucleotides. The target sites are arranged symmetrically relatively to the α-sarcin loop: the sites for oligonucleotides 1, 2, 3 are symmetrical the sites for oligonucleotides 7, 6, 5, respectively. Oligonucleotide 4 is complementary to the entire α-sarcin loop.

Hybridization of the oligonucleotides with ECAS171 RNA was investigated by gel mobility-shift assay. Annealing of oligonucleotides with ECAS171 RNA yielded the same free/complexed RNA ratios as incubation at 37°C for 1 h, indicating that equilibrium is achieved by that time.

Binding of oligonucleotides to the RNA was measured under conditions where the target RNA retain higher order structure. The binding reactions were performed using constant amount of target [^{32}P]-uniformly labeled ECAS171 RNA. Oligonucleotides were taken in concentration ranging from 0.01 μM to 10 μM. Binding of oligonucleotides 1 and 2 to the ECAS171 RNA was too low at the tested oligonucleotides concentrations and not detected by gel mobility-shift assay. For oligonucleotide 3 it was possible only to estimate binding constant in equilibrium binding experiment at 10 μM oligonucleotide concentration.

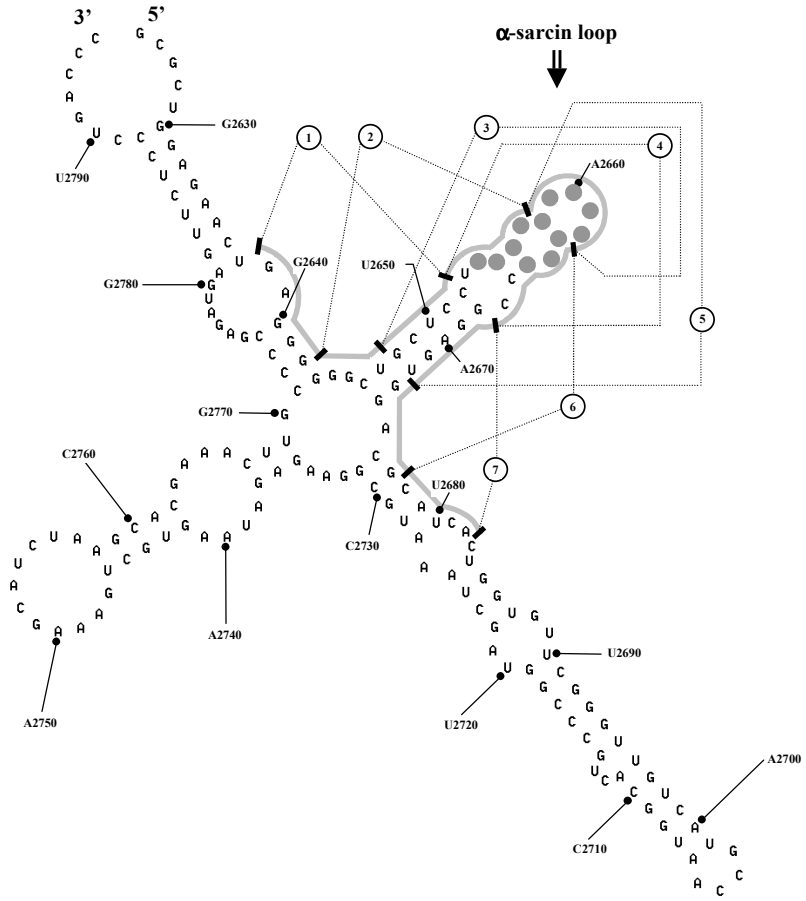


Figure 1. Secondary structure of the ECAS171 RNA fragment of E.coli 23S rRNA and complementary sites for oligonucleotides 1-7. Grey line shows the target region. Marks indicate division of the target region into 5-nucleotide fragments. Complementary sites for each oligonucleotide, consisting of 3 neighboring 5-nucleotide fragments, are marked with dotted lines. 12 the invariable nucleotides of □-sarcine loop are indicated with grey circles.

Discussion

Affinities of the tested pentadecamers to the ECAS171 RNA varied in a wide range, e.g. equilibrium binding constants for oligonucleotides 1 and 2 are less then $<10^4$ while for the oligonucleotide 5 the constant is 4.3×10^7 M^{-1}. So significant differences in affinities can reflect both, stability of the formed heteroduplexes and stability of higher order RNA structure at the oligonucleotide binding site. Moreover, it is quite possible that the part of RNA molecule which is not directly involved in heteroduplex formation can change its conformation due to oligonucleotide binding. We attempted to compare the free energy ΔG parameters for oligonucleotide binding, measured by gel mobility-shift assay, with heteroduplex stabilities [Sugimoto, 1995] and ΔG deduced using several theoretical approaches [Matheus, 1999]. The data are presented in Fig. 2. The obtained experimental ΔG values for oligonucleotides binding do not correlate with calculated heteroduplex stabilities (Fig. 2A, B), evidencing for the crucial dependence of the oligonucleotides affinities to the RNA on the target structure. Figs. 2C, D present ΔG values for the oligonucleotides, predicted by approaches using "RNAstructure 3.5" program. Binding affinities calculated using heteroduplex stability and local structure unfolding parameters are shown in Fig. 2C. The data shown in Fig. 2D were obtained, taking into account the RNA structure refolding caused by oligonucleotide binding. The approach used in the case D, fits the experimentally obtained ΔG values (r^2 = 0.65, P = 0.95). In general, calculated ΔG values are lower than the measured ones: from 0.8 kcal/mole for oligonucleotide 7 up to 8.3 kcal/mole for oligonucleotide 4. This effect can be ascribed to underestimation of RNA structure stability and/or buffer effects. Also, predicted affinity for oligonucleotides 1 and 4 appear to be overestimated compared to relative average binding. By contrast, for oligonucleotide 7, the calculated value appears to be lower then the measured one. Moreover, according to the algorithm used in the case D, oligonucleotide 4 should be the optimal pentadecamer for hybridization with ECAS171 RNA. Our data evidence that, oligonucleotide 5 binds to the target better then oligonucleotide 4. This fact can be ascribed to high stability of the α-sarcin loop structure resulting in underestimation of barrier of the RNA structure unfolding for the oligonucleotide 4. The apparent affinity of oligonucleotide 1 is abnormally lower then predicted (D). The complementary site for oligonucleotide 1 comprises a track of six guanosine nucleotides (4 GC pairs and 2 unpaired G), which possibly can form a stable local structure.

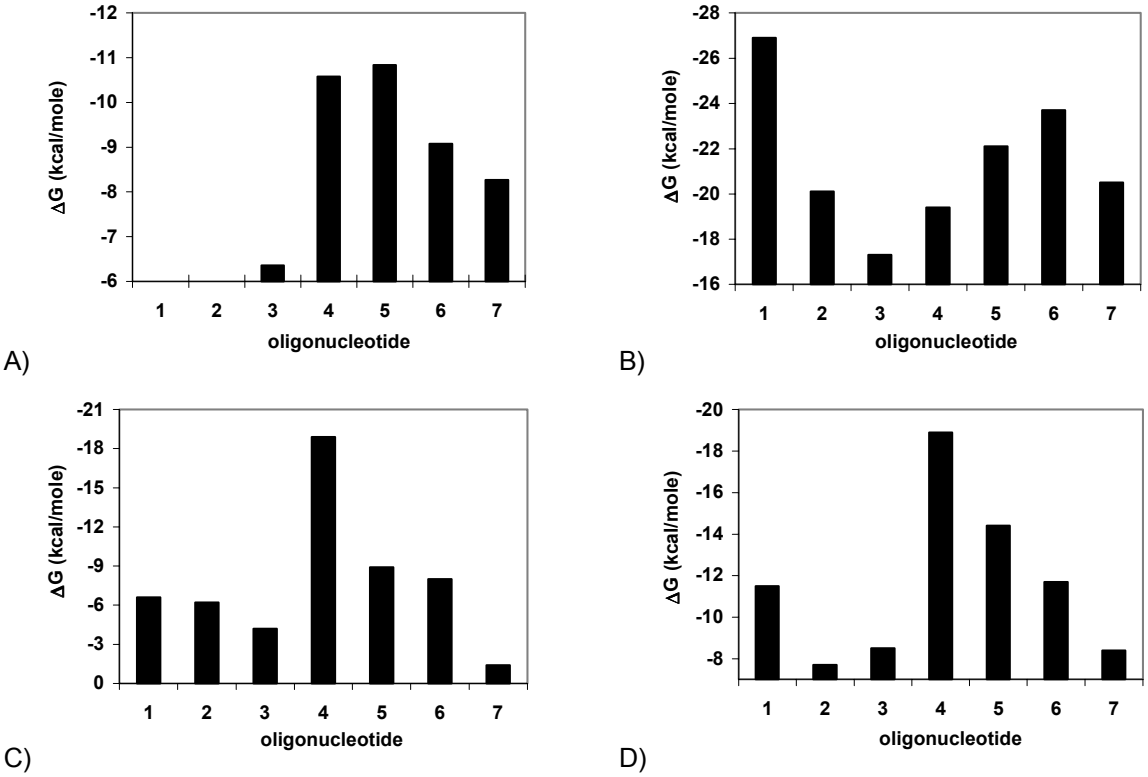


Figure 2. A) ΔG values for oligonucleotides hybridization with ECAS171 RNA at 37°C, obtained using gel-mobility shift assay. B) ΔG values for oligonucleotide heteroduplexes with complementary 15-mer oligoribonucleotide, calculated using "nearest neighbor" approach, for 37°C. C) ΔG values for oligonucleotides hybridization with ECAS171 RNA at 37°C, calculated as ΔG heteroduplex values subtracting ΔG of local RNA structure disruption at oligonucleotide binding site. D) ΔG values for oligonucleotides 1-7 hybridization with ECAS171 RNA at 37°C, calculated as in case C), and taking into account ΔG of RNA refolding after oligonucleotide binding

Conclusions

According to gel mobility-shift data, the oligonucleotide 5, complementary to the 2658-2672 nt, is the optimal pentadecamer for targeting 23S rRNA. Results of the present study demonstrate that currently available theoretical approaches cannot provide an accurate information concerning oligonucleotide affinities even for medium size RNAs. For development of efficient theoretical approaches, contribution of specific stable elements of spatial RNA structure – tetraloops, sites of coordination of magnesium ions and probably G-tracks to the RNA stability should be taken into account.

Acknowledgements

The study is supported by the grant INTAS Nos. 96-1418 and 95-0653 and grant for Young Scientists of SB RAS.

References

1. Branch,A.,D. (1998) A good antisense molecule is hard to find. *Trends Biochem. Sci.*, 23(2), 45-50
2. Correll,C.,C., Wool,I.,G., Munishkin,A. (1999) The two faces of the Escherichia coli 23 S rRNA sarcin/ricin domain: the structure at 1.11 A resolution. *J. Mol. Biol.*, 292(2), 275-287.
3. Endo,Y., Wool,I.,G. (1982) The site of action of alpha-sarcin on eukaryotic ribosomes. The sequence at the alpha-sarcin cleavage site in 28 S ribosomal ribonucleic acid. *J. Biol. Chem.*, 257, 9054-9060.
4. Mathews,D.,H., Burkard,M.,E., Freier,S.,M., Wyatt,J.,R., Turner,D.,H. (1999) Predicting oligonucleotide affinity to nucleic acid targets. *RNA*, 5, 1458-1469.
5. Petyuk,V.,A., Zenkova,M.,A., Giege,R., Vlassov,V.,V. (1999) Hybridization of antisense oligonucleotides with the 3' part of $tRNA^{Phe}$. *FEBS Lett.*, 444, 217-221.
6. Sugimoto,N., Nakano,S., Katoh,M., Matsumura,A., Nakamuta,H., Ohmichi,T., Yoneyama,M., Sasaki,M. (1995) Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, 34(35), 11211-11216.
7. Szewczak,A.,A., Moore.P.,B., Chang.Y.,L. and Wool.,I. (1993) The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc. Natl. Acad. Sci.*, 90, 9581-9585.

DISTRIBUTION OF HISTONES WITHIN DEFINED BAND-INTERBAND REGION OF *DROSOPHILA MELANOGASTER* POLYTENE CHROMOSOMES

Demakov S.A., Schwartz Yu.B., [1]Saumweber H., Semeshin V.F., Zhimulev I.F.

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: demakova@bionet.nsc.ru
[1]Abt. Zytogenetik, Institut fur Biologie, Berlin, Germany
*Corresponding author

Keywords: polytene chromosome, interband, nucleosomes, histone H1, formaldehyde crosslinking, immunoprecipitation

Resume

Motivation:

Despite the large volume of the data on interphase chromosome organization, molecular mechanisms of polytene chromosome banding pattern formation and maintenance still remain unclear. One might propose that chromosome compact (bands) and decompact (interbands and puffs) regions differ in respect of histone distribution. Thus it seemed important to perform the direct molecular and biochemical analysis of the band and interband chromatin in defined chromosome region. We used two approaches. The first one includes the Micrococcal Nuclease (MNase) digestion assay in order to test wether the interband decondensed state is accompanied by the absence of nucleosomes from the interband chromatin. The second approach includes interband DNA cloning and localization of the protein binding sites on the DNA by *in vivo* formaldehyde crosslinking and immunoprecipitation. This approach can be applied for identification the molecular limits of bands and interbands and their boundary condition.

Results:

The direct molecular and biochemical analysis of the band and interband chromatin in defined chromosome regions was performed. The following data were obtained:

1. The 61C7/C8 interband chromatin is organized into nucleosomes in the majority of larval tissues and in polytene chromosomes of salivary glands.
2. Histone H1 is present in the entire 61C7/C8 interband region.
3. H1 relative representation varies threefold within 61C7/8 band/interband region following the banding pattern.

Introduction

The indirect immunofluorescence analysis of antibody binding in general indicates that the distribution of histones corresponds to the banding pattern. Such distribution might appear as a consequence of either high content of the DNA in bands or due to interbands chromatin modifications or because of actual absence of histones from the interband chromatin. In these cases the interbands may represent the stretches of DNA which lack the normal nucleosome packaging and histone H1. Existance of such DNA fragments was shown earlier [Solomon et al., 1988; Nacheva et al., 1989]. However some data [Hill et al., 1989] argue against these findings.

The obvious difficulties in unambiguous interpretation of the cytological data make it of a great importance to perform the direct molecular and biochemical analysis of the band and interband chromatin. It was shown earlier that the P-transposon insertion in the 61C7/C8 region results in formation of a new thin band, which implies that the insertion is within an interband. Using the transposon DNA as a probe the interband DNA was cloned and sequenced and a comparison of the morphological data with the molecular map of the region was performed [Demakov et al.,1993]. In the present work we used this interband region with the aim to investigate the possible differences in DNA-histone interactions between bands and interbands. Two scientific objectives were pursued here:

1. Apply the Micrococcal Nuclease digestion assay in order to test if the interband decondensed state is accompanied by the absence of nucleosomes from the interband chromatin.
2. The molecular mapping of protein/DNA interactions within defined chromosomal intervals. Attention was focused on mapping of histone H1 within 61C7/8 band/interband region following DNA/protein crosslinking and immunoprecipitation.

Results and Discussion

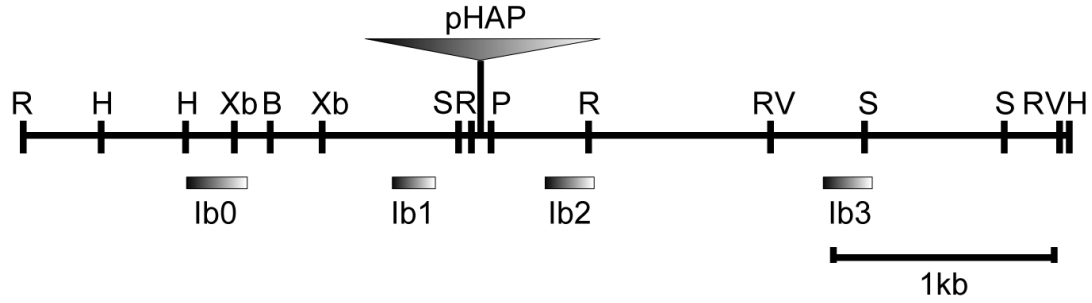


Figure 1. Molecular organization of the 61C7/C8 interband region. The triangle marks the position of a P-transposon insertion which targeted the interband DNA cloning (Demakov et al. 1993). Reference restriction sites within the region are marked with: R-EcoRI; H-HindIII, B-BamHI; Xb-XbaI; S-SalGI; P-PstI and RV-EcoRV. The positions of the interband specific probes are indicated by shaded boxes.

1. Analysis of nucleosomal organization within 61C7/C8 interband region. In order to investigate the ability of the interband DNA to form nucleosomes we have isolated nuclei from the third instar larvae and treated them with MNase. The DNA from such assay was subjected to Southern blot analysis. We used the set of short fragments (about 200 bp) overlapping approximately 3kb of DNA from the 61C7/C8 region as the interband specific probes (Fig. 1). Molecular characterization of 61C7/C8 interband region was made earlier and described elsewhere [Demakov et al. 1993]. 183bp fragment from the coding part of *vermilion* gene was used as the band specific probe. It has been shown previously that this locus is situated in the 10A1-2 band of *Drosophila* X chromosome [Kozlova et al. 1994]. The 10A1-2 band DNA compaction ratio was estimated by Kozlova and co-authors [Kozlova et al. 1994] to vary from 151 to 161. It indicates that the band DNA should form nucleosomes. If the differences in the band and interband compaction correlate with a respective chromatin nucleosome organization one would expect to detect the typical nucleosome array after the band

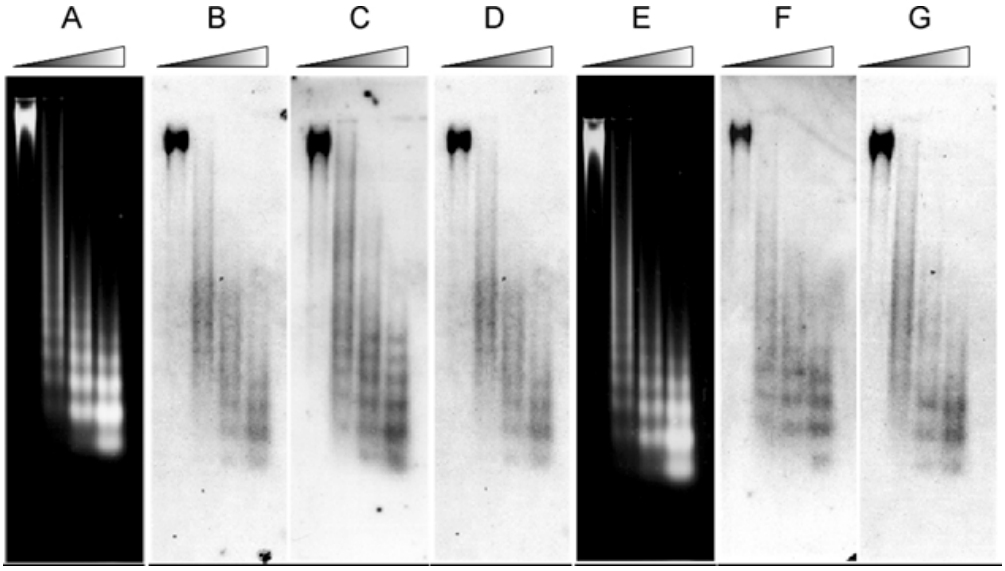


Figure 2. Comparison of the general MNase digestion patterns with the patterns from band and interband sequences in the cells of third instar larvae. The relative enzyme concentration is depicted by shaded triangle above each pannel. A,E. – The images of ethidium bromide stained gels. Southern blot of gel A was probed with: B – 183bp *vermilion* fragment; C – Ib0; D – Ib1. Southern blot of gel E was probed with F – Ib2; G – Ib3.

chromatin MNase digestion and a smear or the irregular pattern at the interband sites.

Figure 2 shows a comparison of the general digestion patterns (Fig. 2A, 2E) and digestion patterns for the chromatin sites homologous to interband specific probes (Fig. 2C, 2D, 2F, 2G). In a general case the typical set of nucleosome oligomers is clearly seen up to the hexamer. Control hybridization with the band specific probe reveals the same nucleosome array (Fig. 2B). Since the numerous cases of tissue specific banding pattern

variation exist, we further addressed whether the interband DNA forms nucleosomes particularly in salivary gland nuclei. Whole salivary glands from third instar larvae were treated with increasing amounts of MNase, the DNA was isolated and analysed as above. The experiments confirm that interband DNA of salivary gland polytene chromosome is also organized into nucleosomes.

Thus we may conclude that the interband chromatin is organized into nucleosomes in the majority of larval tissues and in polytene chromosomes of salivary glands. These data argue against the idea that the interband decondensed state is caused by an inability of the interband DNA *per se* to form the stable nucleosomes. At our knowledge this work represents the first direct investigation of the interband chromatin. The finding of the nucleosomes in the interband region strongly suggests that the difference in compaction state of polytene chromosome bands and interbands appears at the higher levels of chromatin organization.

Whatever might be the reason for the interbands to lack higher levels of chromatin organization (30 nm fiber, for instance) this should correlate with the substantial depletion of H1-histone from the interband chromatin. The latter could be tested experimentally.

2. Mapping of histone H1 within 61C7/8 band/interband region by DNA/protein crosslinking and immunoprecipitation. Histone H1 is involved in the mechanisms of DNA packaging in chromosomes (30 nm fiber). For these reasons this protein has been chosen for our studies of chromatin organization in 61C7/C8 band/interband region. We have shown that monoclonal antibody directed against H1 stained 61C7 and 61C8 bands while no signals have been detected in the interband. This argues that histone H1 is largely reduced or absent in 61C7/C8 interband. To test this possibility the analysis of H1-distribution by formaldehyde crosslinking of DNA/protein complexes *in vivo* and immunoprecipitation have been performed.

The experimental scheme of crosslinking and immunoprecipitation is described elsewhere [Orlando et al.,1997] but the conditions for crosslinking and H1 immunoprecipitation (IP) were optimised.

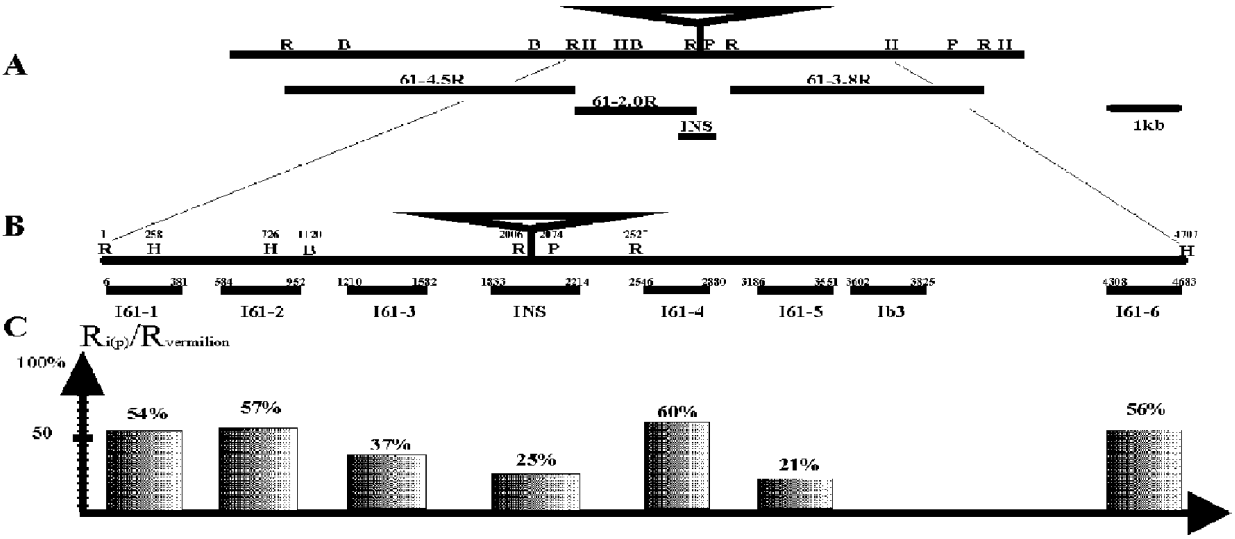


Figure 3. The distribution profile of H1 within 61C7/C8 band/interband region as assayed by IP method. (A) Molecular map of the region. Transposon insertion site into interband region is marked as a triangle. The cloned DNA fragments from this region are shown below. B,H,P,R correspond to BamHI, HindIII, PstI, EcoRI restriction sites. (B) Localization of PCR products used for slot-blot hybridization assay. (C) Hybridization ratio Ri/Rvermilion corresponds to the H1 representation level for the DNA fragments tested.

The specificity and efficiency of the IP have been tested by hybridizing the DNA probe from the *vermilion* locus (band 10A1-2) to the IP-DNA in slot-blot experiments. The profile of H1 distribution within 61C7/C8 region was determined as follows. Hybridization signals were quantified using a BandLeader program and for each membrane the reference value (R_o) was determined by comparing the intensity of a band–specific probe (the 1,8 kb fragment from the coding part of *vermilion* gene) hybridization to the IP-DNA and to the total genomic DNA. Since *vermilion* DNA compaction ratio varies from 151 to 161 [Kozlova et al, 1994], it means that this DNA should be organized as a 30-nm fibril and, therefore, should contain H1. The probe was then stripped off and the membrane was rehybridized up to 3 times with specific probes from 61C7/C8 interband region and the relative enrichment for each particular probe (R_i) was determined. By calculating the ratios R_i/R_o it is possible to draw the distribution profile of H1 within the region of interest. The numerical values of these ratios determined for the 61C7/C8 region are represented on the diagram (Fig. 3).

According to the data obtained several conclusions could be drawn. First, H1 is found over the entire lenght of 61C7/8 band/interband region. These findings indicate that interband chromatin decondensation can occur without complete loss of H1. Second, H1 relative representation within a given region follows the banding pattern. We estimate the interband DNA to be about 3 kb long, situated between DNA fragments including I61-

1, I61-2 probes on the one side and I61-6 on the other one (Fig.3). These fragments are three times enreached in H1 then the internal fragment and thus we belive correspond to the 61C7 and 61C8 bands, respectively. The middle part of the studied region also contains H1 enreached fragment (Fig.3, I61-4 probe). The latter is consistent with the presence of very thin compact band-like material, which could be only seen in some preparations under the electron microscope.

It should be outlined here that in band corresponding fragments the H1 relative enrichment never comprises 100%. These findings might be a consequence of a higher compaction of the thick 10A1-2 band compairing to the thin 61C7 and 61C8 bands. The differences in thick and thin bands compaction have been shown earlier [Kozlova et al., 1994].

Acknowledgements

This work was supported by grant from the INTAS Program (97-1313).

References

1. S.A. Demakov, V.F. Semeshin and I.F. Zhimulev, "Cloning and molecular genetic analysis of *Drosophila melanogaster* interband DNA" Mol Gen Genet. 238, 437-443(1993)
2. R.J. Hill, F. Watt, C.M. Wilson, T. Fifis, P.A. Underwood, G. Tribbick, H.M. Geysen and J.O. Thomas, "Bands, interbands and puffs in native *Drosophila* polytene chromosomes are recognized by a monoclonal antibody to an epitope in the carboxy-terminal tail of histone H1" Chromosoma 98, 411-421(1989)
3. T.Yu. Kozlova, V.F. Semeshin, I.V. Tretyakova, E.B. Kokoza, V. Pirrotta, V.E. Grafodatskaya, E.S. Belyaeva and I.F. Zhimulev, "Molecular and cytogenetical characterization of the 10A1-2 band and adjoining region in the *Drosophila melanogaster* polytene X chromosome" Genetics 136, 1063-1073(1994)
4. G.A. Nacheva, D.Y. Guschin, O.V. Preobrazhenskaya, V.L. Karpov, K.K. Ebralidse and A.D. Mirzabekov, "Change in the pattern of histone binding to DNA upon transcriptional activation" Cell 58, 27-36(1989)
5. V. Orlando, H. Strutt and R. Paro, "Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. Methods" 11, 205-214(1997)
6. M.J. Solomon, PL. Larsen and A .Varshavsky, "Mapping protein–DNA interactions *in vivo* with formaldehyde: Evidence that Histone H4 is retained on a highly transcribed gene" Cell 53, 937-947(1988)

DETECTION OF LOCATION AND TYPE OF SINGLE NUCLEOITIDE REPLACEMENT IN ANY AMPLIFICATED REGION OF GENOME. COLORIMERTRIC TEST SYSTEMS BASED ON LIGATION OF TANDEM OF SHORT OLIGONUCLEOTIDES

Dymshits G.M., Kabilov M.R., Neischtadt V.P., Skobeltsyna L.M.** [1]**Zarytova V.F.,** [1]**Ivanova E.M.,** [1]**Pyshnyi D.V.

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: dimshits@niboch.nsc.ru
[1]Novosibirsk Institute of Bioorganic Chemistry, Novosibirsk, Russia
e-mail: pyshnyi@niboch.nsc.ru
*Corresponding author

Keywords: point mutation, DNA diagnoctics, mismatch, oligonucleotides, ligation, polymer support

Resume

Several approaches are currently used to detect point mutations in PCR-amplified fragments of the genomic DNA. The methods used most widely are based on restriction analysis (PCR-RFLP), denaturing gradient gel-electrophoresis (DGGE), allele-specific polymerase chain reaction (PASA, ARMS), direct hybridization with oligonucleotide probes. All these methods are not universal and are used to detect certain mutations. Restriction analysis methods may be applied to detect mutations in restriction sites only. DGGE requires special high-cost equipment. Hybridization of oligonucleotide probes with target DNA is the base of the microchip technology. This reversed hybridization method is now extensively developed as a universal approach to study of DNA sequence and to detect the point mutations. The approaches based on hybridization with oligonucleotide probes require precise conditions of analysis because of low sensitivity of the extended oligonucleotide probes to single-nucleotide substitutions in the target DNA and essential differences in stability of complexes formed oligonucleotides with identical length in case of microchip technology. Reliability of the obtained result with the use of oligonucleotide probes may be improved *via* additional reactions with participation of specific enzymes sensitive to mismatches in the complexes of oligonucleotide and DNA template. The enzymes may be DNA- and RNA-polymerases (PASA, ARMS), or DNA ligases. DNA ligase allows one to detect single-nucleotide substitutions in DNA because ligation of two oligonucleotides becomes much less efficient if a non-complementary base pair of the complex formed with the DNA template and oligonucleotides is located in close proximity to the ligation site. Selectivity of DNA ligase depends on the type of the single-nucleotide substitution and on its location related to the phosphodiester bond whish is to be formed between oligonuclotides. However, in case of extended probes substrate specificity of enzyme cannot provide the ligation exclusively in perfect complexes of probes with DNA. One of the approaches to improve the ligation method for detection of single-nucleotide substitutions in the sequence of DNA template is shortening of one of the two ligating oligonucleotide probes. This is obviously related to low hybridization properties of short (6-10-mer) oligonucleotides: the complexes with at least one non-complementary base pair are notably less stable. However, even in these conditions a single-base substitution never resulted in complete absence of the ligation product. As shown in our earlier work [1] formation of a «wrong» product may be practically totally excluded when three short oligonucleotides in tandem: octamer-tetramer-octamer are ligated on DNA template and mismatch is located in the complex of central tetranucleotide. High ligation specificity of these tandems is provided by high selectivity of tetranucleotide binding to the template in the presence of the flanking octamers and by enhanced selectivity of the enzyme action, because of any unpaired base of the tetranucleotide duplex in $pN_8+pN_4+pN_8$ tandem is close to the ligation sites.

Test systems for detection of point mutations at diagnostic of hereditary diseases should identify the signal products with non-radioactive reporter groups. The system including biotin plus streptavidin-alkaline phosphatase conjugate that allows one to detect 10^{-14} mole DNA is used rather often. Colorimetrtic detection of the ligation product containing a biotin residue with using streptavidin-alkaline phosphatase conjugate is possible only after removal of remaining initial biotinylated short oligonucleotide from the final reaction mixture. One of the ways to solve this problem is ligation of oligonucleotides one of which is immobilized on the solid-phase support.

In this work we have studied colorimetric detection of the product of ligation of the oligonucleotide tandem composed of octamer-tetramer-octamer, where one octamer is bound by a 12-members amino linker to the polymeric beads DMEG (**B~pN_8**) and the other contains the biotin residue (**pN_8Bio**); the 20-mer oligonucleotide M was used as a DNA template (Fig. 1).

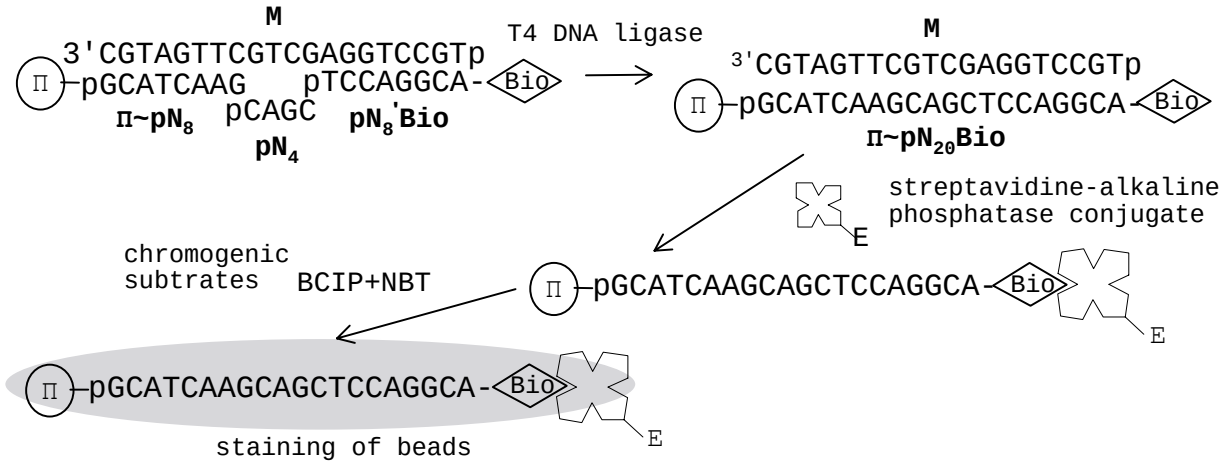
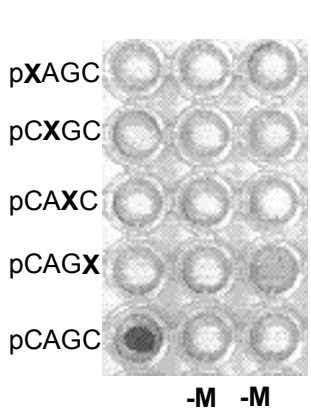


Figure 1. Colorimetric detection of the product of ligation.

Ligation of oligonucleotides **B~pN$_8$+pN$_4$+pN$_8$'Bio** on the complementary template M was performed at 37°C for 60 min, then the polymer carrying the ligation product **B~pN$_{20}$Bio** was washed and treated with streptavidin-alkaline phosphatase conjugate for 30 min. After washing off the conjugate not bound with the biotin residue of the ligation product the «development» reaction was carried out in presence of chromogenic substrates BCIP and NBT. The oxydation-reduction reaction of these substartes is initiated by phosphatase and produces water-insoluble colored products.

The efficiency of test system for detection of single-base mismatches in the tetramer duplex was studied in model systems with all possible substituted tetramer sequences. Thus we performed ligation of the 13 systems 12 of which had mismatches in tetramer complexes and the template M.

As seen from Fig. 2, intense staining of polymer beads is observed only for the tetranucleotide **pCAGC**, which is fully complementary to the DNA template. The ligation of all tandems containing tetramers with a single-base substitution results in staining with lowest intensity. Therefore, any single-base substitution in the tetranucleotide component of the tandem strongly inhibits ligation of the tandem [1] and is reliably detected by staining of polymer beads with the biotin-streptavidin technique.



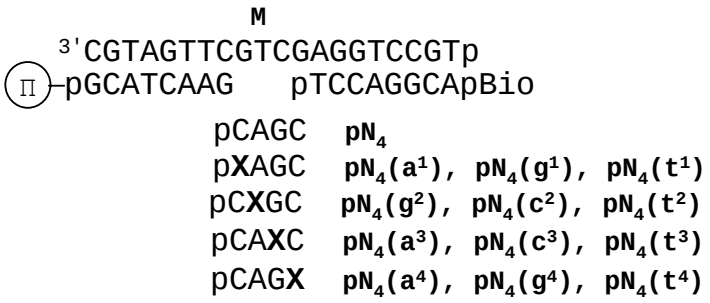


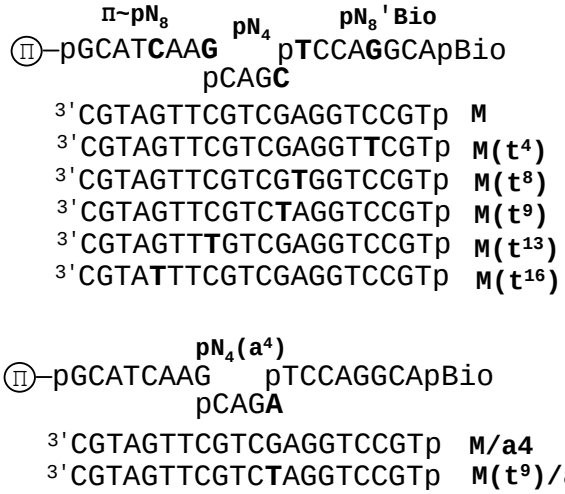
Fig. 2. Scanned image of the panel after 20 min development after ligation of tandems in perfect or imperfect complexes on templates M and treatment with streptavidin-alkaline phosphatase conjugate and with substrates; [M]=10^{-7} M.

Efficiency of the test system toward point mutations in the model DNA templates was studied introducing single-base substitutions in either tetramer-binding or octamer-binding sites of the 20-mer DNA template. "Mutations" in the octamer-binding sites were either at the ligation point or rather far from it (Fig. 3).

It is evident that the template M(t9) with a single-base substitution in the tetramer-binding site is strongly discriminated at ligation of the tandem **B~pN$_8$+pN$_4$+pN$_8$'Bio**, whereas in the system including Mt9 template with

the complementary tandem **B~pN$_8$+pN$_4$(a^4)+pN$_8$'Bio** staining was not weaker than in the "perfect" system **M+(B~pN$_8$+pN$_4$+pN$_8$'Bio)**. The templates M(t4) and M(t8) with single-base substitutions in the biotinylated octanucleotide **pN$_8$'Bio**-binding site are practically not discriminated. On the contrary, ligation of the tandem **B~pN$_8$+pN$_4$+pN$_8$'Bio** on the templates M(t13) and M(t16) carrying substitutions in the site of binding of the immobilized octanucleotide B~pN$_8$ produces no signal product.

The similar test-system was developed on the basis of activated lavsan. This membrane has a number of dignities doing it perspective to microchip technology development of DNA diagnostics: it is chemically inert, flexible, transparent, with a homogeneous not moistened surface.



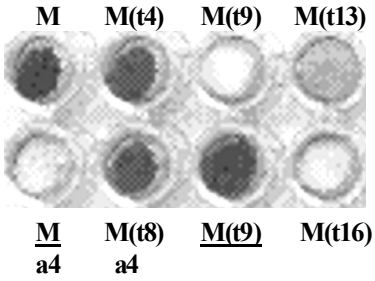


Fig.3. Scanned image of the panel after development after ligation of tandems in perfect or imperfect complexes on templates M and M(tn) and treatment with streptavidin-alkaline phosphatase conjugate and with substrates.

The third test - system is based on ligation of the tandem **pN$_8$+pN$_4$+pN'$_8$Bio** on a DNA template in solution. In this case the method of UV immobilization of extended DNA moleculs on nylon [2] was used to separate the biotinylated product of ligation **pN$_{20}$Bio** and initial octamer **pN'$_8$Bio**. The UV irradiation of ligation mixture results in immobilization only analyzed DNA due to the large difference in molecular weights. The immobilizated DNA template forming with 20-mer product **pN$_{20}$Bio** stable complementary complex keeps it on membrane while the octamer **pN'$_8$Bio** having weaker hybridization properties is washed up easily (Fig. 4).

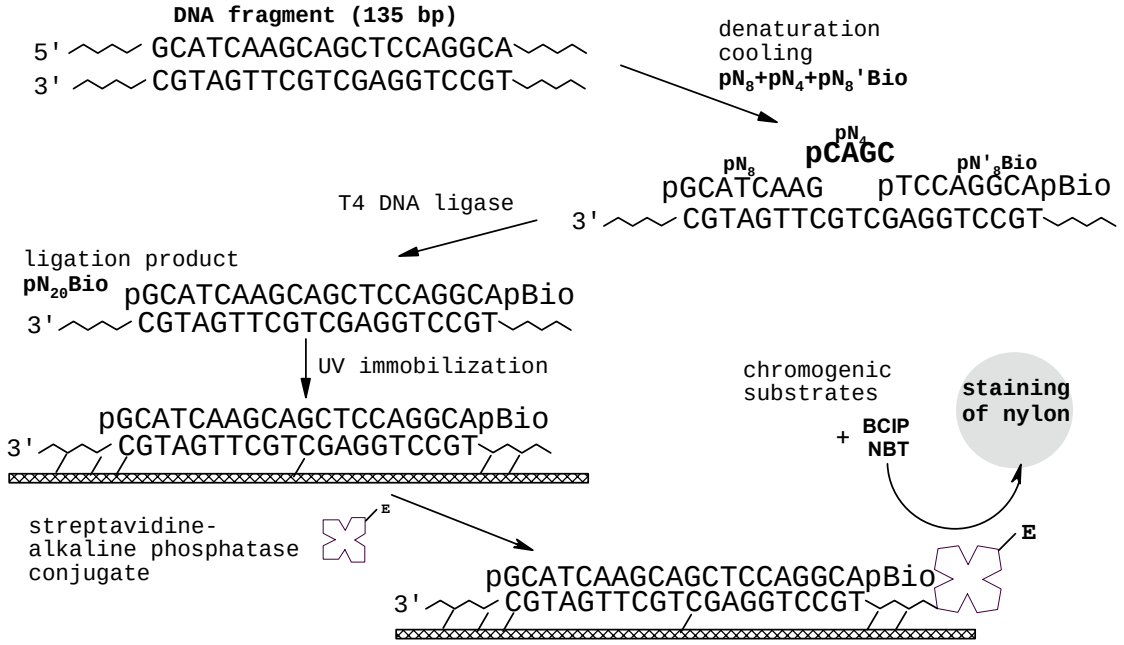


Figure 4. The immobilizated DNA template forms with 20-mer product pN20Bio stable complementary complex keeps it on membrane.

The possibility of revealing of point mutation in genomic DNA was studied using DNA fragment HIV-I (135 bp) and tandems containing tetramers with all possible single nucleotide substitution. The ligation of 13 tandems were performed, 12 of which had mismatches in tetramer complexes and the template DNA. The intensive coloring of spot on membrane was observed only in the case of ligation of tandem containing tetramer pCAGC

which was fully complementary to seguence of analyzed DNA (Fig. 5). The "development" of ligation product when other tandems were used caused the lowest staining of membrane though the their coloring depended on mismatched type.

Thus we suggested colorimetric test systems which allowed to detect unambiguously and with highly sensitive either location and type of single nucleotide replacement in any amplificated region of genomic DNA during 2-3h.

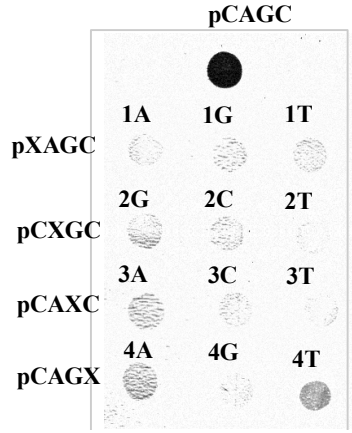


Fig. 5. Scanned image of nylon after "development" of the products of ligation hybridizating with UV immobilized DNA fragment.

This work was made possible by Grants from Russian State Program "The new methods of bioengineering" ("Human Genome") and INTAS - 96-1778.

References

1. Pyshnyi D.V., Krivenko A.A., Lokhov S.G., Ivanova E.M., Dymshits G.M., Zarytova V.F. Interaction of derivatives of short oligonucleotides with nucleic acids: VI. Discrimination of mismatch-containing complexes upon ligation of a tandem of short oligonucleotides on DNA template. Bioorg. Khim. (Rus.) 1998. V. 24. P. 32-37.
2. Kalachikov S.M, Adarichev B.A., Dymshits G.M. Immobilization of DNA on microporous membranes using UV-irradiation. Bioorg. Khim. (Rus.) 1992. V. 18. P. 52-62.

XENOTRANSPLANTATION OF DROSOPHILA EMBRYONIC NERVE CELLS INTO MAMMALS BRAIN: MOLECULAR GENETIC AND CLINICAL ASPECTS

[1,3]Korochkin L.I., [3]Alexandrova M.A., [1]Bashkirov V.N., [1]Dzitoeva S.G., [1]Pavlova G.V., [3]Trukhatcheva A.A., [2]Evgeniev M.B., [1]Modestova E.R., [4]Grosman A.A.

[1]Institute of Gene Biology, Moscow, Russia
e-mail: likor@mx.ibg.relarn.ru
[2]Institute of Molecular Biology, Moscow, Russia
[3]Institute of Developmental Biology, Moscow, Russia
[4]In Vitro Diagnostics, Inc, New York, USA
*Corresponding author

Resume

Transgenic Drosophila melanogaster stocks bearing foreign genes (lacZ, gdnf, bdnf, ngf) were selected. Constructs based on the CaSper vector, containing bacterial gene lacZ or other foreign genes, were made and microinjected into Drosophila embryos. Modern technique allows localization of Drosophila cells transplanted to a foreign brain by X-gal staining. The constructs containing genes coding for GDNF, BDNF, NGF, were made and introduced into Drosophila genome. It was shown that embryonic nerve cells from Drosophila neuromutants survived in mammalian brains (rats, mice) for 10-12 days. Subsequently they were attacked by macrophages. A combination of homografts and xenografts of embryonic nerve tissue were transplanted into rat brain. Drosophila cells prevented formation of glial scar and stimulated the vascularization of the graft area. The presence of Drosophila cells promotes survival and differentiation of homograft cells. This method has been successfully applied to treatment of the Parkinson's disease in humans.

The study is supported by the grant INTAS No. 96-1439.

REDOX MODULATION OF DNA TOPOISOMERASE I ACTIVITY IN PLANTS

Konstantinov Y.M., Tarasenko V.I., [1]Rogosin I.B.

Institute of Plant Physiology and Biochemistry SB RAS, Irkutsk, Russia
[1]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: yukon@sifibr.irk.ru
*Corresponding author

Keywords: DNA topoisomerases I, gene, core domain, conservative cysteines, redox modulation, evolution

Resume

DNA topoisomerases along with DNA and RNA polymerases act as key enzymes of genetic information transfer, the cell deprived of topoisomerases fails to make up for their absence and thus perishes. These enzymes participate in many fundamental genetic processes associated with separation of DNA strands such as replication, transcription, recombination and repair. Topoisomerases may also act as DNA strand transferases and catalyze recombination and transposition reactions [Pommier, 1998]. Type I DNA topoisomerase act by making a transient nick on a single-strand of duplex DNA, passing another strand through the nick, and changing the linking number by one unit [Gupta et al.,1995]. Beyond their normal physiological functions, eukaryotic topoisomerases I have been identified as the primary cellular targets for a variety of antitumor agents [Pommier, 1998].

Plant cells contain topoisomerases not only in nucleus, but also in mitochondria and chloroplasts. No detailed molecular biological studies of plant topoisomerases I of nuclear and mitochondrial localization have been accomplished up to now - neither on the level of encoding genes, nor on the level of their products. It should be noted that molecular mechanisms of regulation of this enzyme activity remain practically unexplored. These mechanisms apparently differ for genetic systems of nucleus and organelles.

The aim of the present work was molecular biological investigation of DNA topoisomerase I of higher plants to reveal peculiarities of structural organization of its protein molecule and possible biochemical mechanisms of its activity regulation. Model three-dimensional structure of topoisomerase I of carrot developed on the basis of the data of crystalline structure of human topoisomerase I, was analyzed to check the hypothesis [Konstantinov, Tarasenko, 1999] on possible participation of some highly conservative cysteine residues of core domain in redox modulation of the enzyme activity.

Based on the analysis of sequence and functional mapping of eukaryotic topoisomerase I can be divided into four principal domains: a sparsely conserved N-terminal domain, a highly conserved core domain, a poorly conserved linker domain and a highly conserved C-terminal domain including the active site tyrosine (Fig. 1).

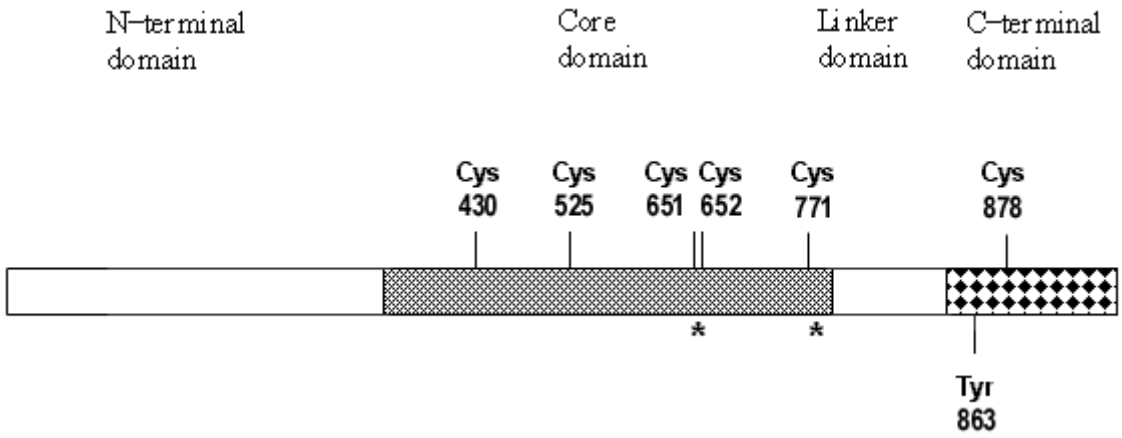


Figure 1. Domain structure of DNA topoisomerase I of carrot. The scheme is developed on the basis of the data of Balestrazzi et al.(1996). Cys – cysteine, Tyr – tyrosine of topoisomerase I active center. The numbers indicate the localization of highly conserved cysteine residues. Asterisk represent sites whose cysteines may form disufide bond according to the computer modeling data.

Taking into account earlier acquired data on the importance of sulfhydryl groups for the activity of DNA topoisomerases of plant origin we hypothesized that protein molecule of this enzyme may contain redox

sensitive regulatory cysteines ensuring sensitivity of topoisomerase I to the change of redox state of glutathione and, possibly, other low molecular biothiols in the cell. Experimental study of redox conditions impact on the activity of DNA topoisomerase I in plants showed that under oxidizing conditions created by addition of potassium ferricyanide or oxidized glutathione (GSSG) significant decrease of topoisomerase activity was observed, whereas under reducing conditions created by addition of sodium dithionite or reduced glutathione (GSH), enzyme activation was observed. GSSG significantly exceeded potassium ferricyanide in the degree of inhibition of DNA relaxation activity of topoisomerase, at the same time GSH and sodium dithionite produced approximately the same activating effect. Impact of redox system of glutathione on topoisomerase activity observed in the course of experiments, in our view, indicates possible participation of reactions of thiol-disulfide exchange on the level of cysteine residues within highly conservative core domain of topoisomerase I in its activity modulation. We conducted comparative analysis of topoisomerase I in eukaryotes. Sequences were aligned by using the MAP program [Huang, 1994]. The topoisomerase I amino acid sequences were taken from the PIR database. A fragment of alignment of four plant topoisomerases I is presented in Fig. 2. It may be seen that the topoisomerase I sequence is very conservative. Analysis of amino acid sequences showed the presence of 6 highly conservative cysteine residues, localized largely in the core domain (Fig. 2). One conservative cysteine is located in the C-terminal domain. Based on the results of the experiments on the impact of oxidized and reduced forms of glutathione on the relaxation activity of mitochondrial topoisomerase of carrot, we hypothesize that a certain pair of these cysteines is liable to accomplish reversible formation of regulatory disulfide bond resulting in the modulation of enzyme activity. For example, cysteines at positions 651-652 are close enough to a cysteine at position 771, and they might form a disulfide bond according to a predicted tertiary structure (results not shown).

Figure 2. Alignment of amino acid sequences of DNA topoisomerase I of different plant species. Only regions containing conservative cysteine residues are shown. Conservative cysteines are highlighted in grey. 1 - *Daucus carota*; 2 - *Nicotiana tabacum*; 3 - *Arabidopsis thaliana*; 4 - *Pisum sativum*.

Phylogenetic analysis of nucleotide sequences was performed by using the computer programs MEGA 1.03 [Kumar, Tamura, and Nei, 1993] and PAUP* 4.0 [Swofford, 1998]. The neighbor-joining and maximum parsimony methods were used to reconstruct phylogenetic trees from nucleotide sequence data. The bootstrap test was conducted with 1000 replications. The proportion of nucleotide sites at which two nucleotide sequences are different (*p*-distance) and corrected Poisson distance were used as measures of the extent of sequence divergence. Figure 3 shows the neighbor-joining tree of topo I sequences. We did not include in the analysis

sequences of topoisomerases I from *Plasmodium falciparum* and *Ustilago maydis* due to significantly accelerated evolution rates of these sequences compared to other investigated topoisomerases. Phylogenetic tree is in a good accordance with a commonly accepted viewpoint on Eukaryotae evolution. Divergence of fungi, plants and animals topo I can not be resolved due to low bootstrap value (54) for the fungi - plant cluster (Fig. 3). Very similar results were obtained with the maximum parsimony analysis (results not shown). The results acquired show that topoisomerase I evolution is characterized by a strong purifying selection at the core

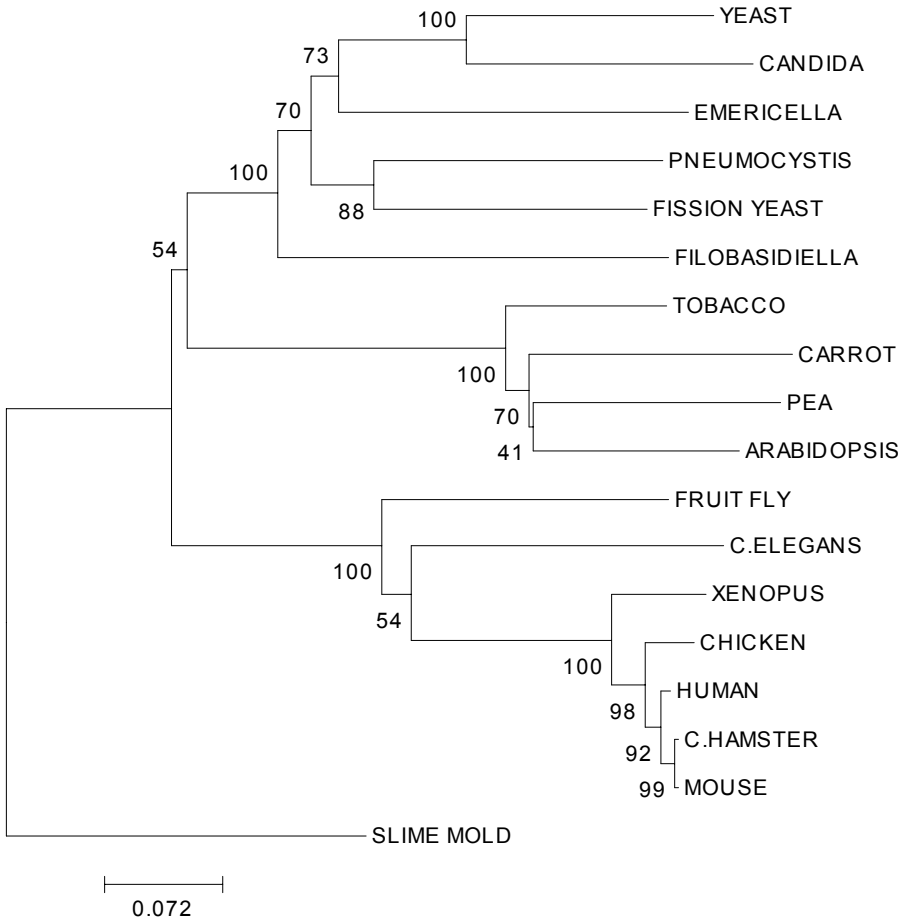


Figure 3. Neighbor-joining tree (Poisson distance) for eukaryotic topoisomerase I amino acid sequences. The numbers for the interior branches refer to the bootstrap values with 100 replicates.

domain and absence of intensive horizontal gene transfer.

Acknowledgments

The financial support of INTAS Grant No 98-0522 is acknowledged.

References

1. Balestarzzi, A., Toscano, I., Bernacchia, G., Luo, M., Otte, S., and D.Carbonera, Cloning of a cDNA encoding DNA topoisomerase I in *Daucus carota* and expression analysis in relation to cell proliferation. Gene **183**, 183 (1996).
2. Gupta, M., Fujimori, A., and Y. Pommier, Eukaryotic DNA topoisomerases I. Biochim. Biophys. Acta **1262**, 1 (1995).
3. Huang, X., On global sequence alignment. Comput. Applic. Biosci. **10**, 227 (1994).
4. Konstantinov, Yu., and V. Tarasenko, Changes of DNA topoisomerase I activity from maize mitochondria under the influence of redox conditions. Maize Genet. Coop. Newslett. **73**, 39 (1999).
5. Kumar, S., K. Tamura, and M. Nei, MEGA: Molecular Evolutionary Genetics Analysis, Version 1.02. The Pennsylvania State University, University Park, PA. (1993).
6. Swofford, D.L., PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Suderland, Mass (1998).
7. Pommier, Y. Diversity of DNA topoisomerases I and inhibitors. Biochimie **80**, 255 (1998).

ARTIFICIAL RIBONUCLEASES: NOVEL TOOLS FOR INVESTIGATION OF RNA STRUCTURE IN SOLUTION

Zenkova M.A., Konevetz D.A., Beloglazova N.G., Chumakova N.A., Vlassov V.V., Sil'nikov V.N.

Institute of Bioorganic Chemistry SB RAS, Novosibirsk, Russia
e-mail: marzen@niboch.nsc.ru
*Corresponding author

Keywords: RNA cleavage, ribonuclease mimics, RNA structure probing

Resume

Small organic molecules mimicking the catalytic centers of ribonucleases A and T1 have been developed [1,2]. One or two imidazole residues, amino and carboxylic groups in different combinations were connected by flexible linkers to cationic structures based on 1,4-diazabicyclo[2.2.2]octane fragments with aliphatic fragment. All the mentioned domains: the imidazole residues, dicationic structure and aliphatic fragment were shown to be essential for the ribonuclease activity of the compounds.

Ribonuclease activity of the designed RNase mimics was investigated in experiments with oligoribonucleotides, tRNAs and viral RNAs. Designed ribonucleases catalyze hydrolysis of RNA substrates yielding 2',3'-cyclophosphate and 5'-OH groups at the cleavage site. The artificial ribonucleases display pronounced structure- and sequence- specificity similar to that of RNase A in wide range of buffers and incubation conditions. Rate of hydrolysis of individual phosphodiester bonds in RNA was strongly affected by RNA secondary and tertiary structure. These mimics, like RNase A, cleave RNA substrates (yeast $tRNA^{Phe}$, yeast $tRNA^{Asp}$ influenza virus M2 mRNA) preferentially at UA and CA sequences located within the single-stranded regions.

Introduction

The compounds capable of cleaving phosphodiester bonds in RNA are of great interest in view of a variety of problems to be solved therewith. Among such problems, the design of probes for investigation the structures of RNA and RNA-protein complexes in solution [3] is to be singled out. As compared to enzymatic probes, small chemical reagents withstand a wider range of conditions and do not perturb the object of study. Probing RNA structure with compounds reacting with phosphodiester linkages is particularly informative, because in these experiments reactivities of universal structures present in all RNA residues can be assayed in one experiment. On the other hand, these compounds, capable of cleaving RNA in physiological conditions, may be useful as antiviral agents.

Artificial RNases can be obtained by conjugation of hydrolytically active groups with molecules possessing an enhanced affinity to nucleic acids. Recently we synthesized artificial RNases [2-4] bearing imidazole residues, which play the key role in hydrolysis of phosphodiester bonds by natural RNases. The enhanced affinity of such constructs to RNA is provided by the presence of positively charged fragments - a bisquaternary salt of diazabicyclo[2.2.2]octane. Here, we report that these compounds can be useful in studying the secondary structure of RNA.

Methods

Native yeast $tRNA^{Phe}$ was a generous gift of Prof. G.Keith, *in vitro* transcript of human mitochondrial $tRNA^{Lys}$ with a point mutation A9→C and influenza virus M2 RNA were synthesized enzymatically *in vitro* using T7 RNA polymerase and plasmids linearized by corresponding restriction enzyme essentially as described earlier.[35]. Bulk yeast tRNA, used as RNA carrier to supplement labeled RNAs, was from Boehringer-Mannheim (Meylan, France).

RNAs were 5'-end labeled by dephosphorylation with alkaline phosphatase followed by phosphorylation with [γ-^{32}P] ATP and polynucleotide kinase as described [5]. To introduce radiolabel at the 3'-end of the RNAs, the labeling was performed by ligation of [5'-^{32}P] pCp with the T4 RNA ligase [2]. After labeling, RNAs were purified by electrophoresis in 12% denaturing polyacrylamide gels. The labeled RNAs were eluted from gels and ethanol precipitated, RNAs were dissolved in water and stored at -20°.

Cleavage Experiments

Cleaving experiments are usually performed with 50,000 to 100,000 Cerenkov cpm of labeled RNA. Treatment of the RNAs with chemical constructs was performed at 37°C for 1-7 h. Reaction mixtures (10 μl) contained 50 mM imidazole buffer, pH 7.0, 200 mM KCl, 0.1mM EDTA, 100 μg/ml tRNA carrier (total tRNA from *Escherichia coli*), and one of the compounds at concentration of $5x10^{-5}$ M - $2x10^{-4}$ M. In some experiments the buffer was 50 mM HEPES-KOH, pH 7.0, or 50 mM cacodylate, pH 7.0, or 50 mM phosphate, pH 7.0 or 50 mM TRIS-HCl pH 7.0. After incubation, reactions were quenched by precipitation of the RNA with 2% lithium perchlorate solution in aceton. The cleavage products were resolved by electrophoresis in 18% polyacrylamide gel containing 8 M urea. The gels were autoradiographed on the Agfa X-ray film. To obtain quantitative data, the oligonucleotide bands were cut out of the gel and radioactivity of the gel slices was determined by Cherenkov's counting.

Results and Discussion

In this communication we describe of RNA cleaving properties of conjugates of 1,4-diazabicyclo[2,2,2]octane with imidazole (histamine) group (D1) and imidazole and carboxylic groups (histidine) (D2 and D3). The conjugates D1 – D3 are composed of three domains: imidazole residue (A), dicationic structure (B) and aliphatic fragment (C) (Fig. 1). Hydrolytic activity of artificial nucleases was studied in experiments with ribooligonucleotide UUCAUGUAAA, 3'-[^{32}P]-$tRNA^{Phe}$, and 5'-[^{32}P]- $tRNA^{Lys}$. Reactions were performed at 37°C in 50 mM Imidazole buffer pH 7.0. It was found that compounds D1 -D3 demonstrate substantial ribonuclease activity: 60, 90 and 100% of $tRNA^{Phe}$ was depolymerized by 18 h incubation by D1, D2 and D3, respectively. Complete cleavage of UUCAUGUAAA, M2 RNA and $tRNA^{Phe}$ by D3 required 4 h, 3 h and 18 h incubation, respectively, indicating that the rate of RNA hydrolysis was strongly affected by the RNA structure. These compounds display similar cleavage specificity with some variations. Sensitivity of phosphodiester bonds to the synthetic ribonucleases decreases in the order: CA>UA>>CG>UG>>CC, UU, GG, AA. The compounds display very strong preference to the single-stranded RNA regions.

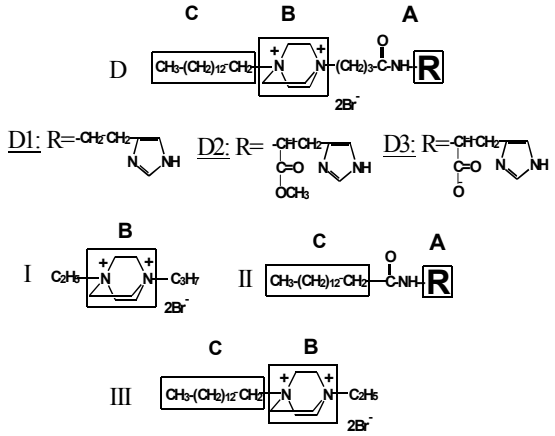


Figure 1. Chemical ribonucleases D1 – D3 a and compounds I – III. The structural domains are indicated by A, B, and C.

Compounds I - III (Fig.1) representing parts of ribonuclease mimics D1 – D3 are composed of diazabicyclooctane moiety (I), aliphatic fragment connected with histamine residue (II), and aliphatic fragment connected with diazabicyclooctan (III). The compounds I and II do not cleave the RNA at all and compound III demonstrates faint hydrolytic activity. According to rationale which was behind the design of the RNase mimics, it was satisfying to observe that the compounds lacking either diazabicyclooctane moiety, or histamine are inactive. Apparently, both catalytic histamine residue and the RNA binding dicationic structure are needed for the catalytic activity.

We have investigated the mechanism of RNA hydrolysis by the chemical ribonucleases. It is shown that the compounds cleave by mechanism similar to that of RNase A, through 2',3'-cyclophosphate as the intermediate product. Comparison of hydrolysis of CpA in water and D_2O revealed the isotope effect K_H/K_D = 2.28), which indicates the acid-base catalysis at the limiting step of the reaction.

Nature of buffer affects the reaction to a limited extent. In HEPES, cacodylate, phosphate, the reaction occurs somewhat less efficiently than in imidazole buffer whereas in Tris buffers at pH 7.0 the reaction occurs with the rate similar to imidazole buffer. The enhanced rate of RNA hydrolysis in imidazole buffer may suggests participation in the process of buffer components, neutral or protonated imidazole species, capable of catalysis themselves. The reaction is inhibited by salt, apparently because increase of ionic strength interferes with electrostatic interaction between cationic D1 with polyanionic substrate RNA.

Investigation of the rate of tRNA hydrolysis versus D3 concentration shows a bell-shaped dependence with maximum at 0.2 mM of D3, perhaps reflecting optimal density of positively charged conjugates on the RNA The cleavage pattern of human mitochondrial $tRNA^{Lys}$ with point mutation A9→C, generated by D3 and RNase A is presented in Fig. 2A. Conjugate D3 generates a non-random hydrolysis pattern similarly to that of RNase A with the major cuts at phosphodiester bonds after A9, C20, U23, U43, C52, and C55 within single-stranded RNA regions and junctions (Fig. 2B). The cleavage pattern of tRNA may be affected by two factors, that are the reactivities of individual phosphodiester bonds and the specificity of binding of the cationic molecules at the tRNA surface. The data indicate that the specificity of the RNA cleavages by conjugates D1 – D3 is mainly governed by the reactivity of individual phosphodiester bonds. The conjugates demonstrate pronounced preference to the pyrimidine-purine sequences in particular, to CpA and UpA. The access of cleaving molecules

to phosphodiester bonds in different dinucleotide sequences is affected by the nature of residues in vicinity of the bond to be cleaved and by conformational and freedom of the linkage. Apparently, a favorable situation is realized for PypPu sequences. The unstability of such sequences in RNA is a known phenomenon, the exact nature of which remains to be elucidated.

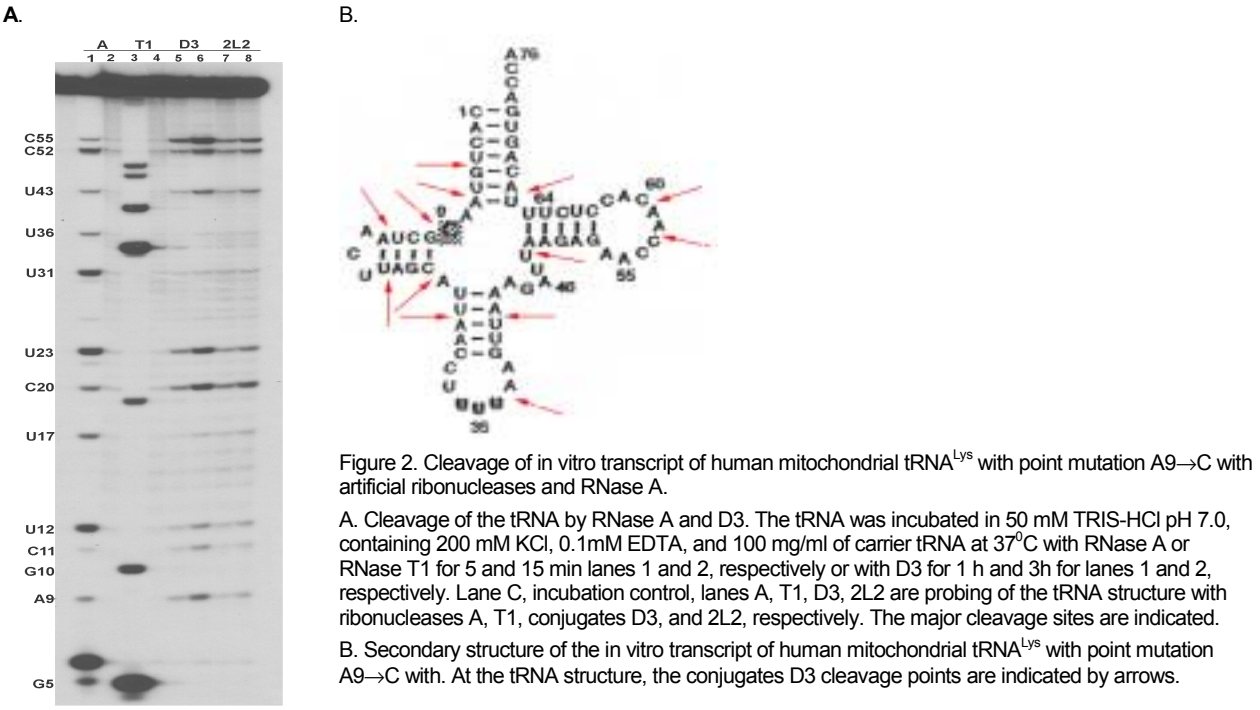


Figure 2. Cleavage of in vitro transcript of human mitochondrial $tRNA^{Lys}$ with point mutation A9→C with artificial ribonucleases and RNase A.

A. Cleavage of the tRNA by RNase A and D3. The tRNA was incubated in 50 mM TRIS-HCl pH 7.0, containing 200 mM KCl, 0.1mM EDTA, and 100 mg/ml of carrier tRNA at 37^0C with RNase A or RNase T1 for 5 and 15 min lanes 1 and 2, respectively or with D3 for 1 h and 3h for lanes 1 and 2, respectively. Lane C, incubation control, lanes A, T1, D3, 2L2 are probing of the tRNA structure with ribonucleases A, T1, conjugates D3, and 2L2, respectively. The major cleavage sites are indicated.

B. Secondary structure of the in vitro transcript of human mitochondrial $tRNA^{Lys}$ with point mutation A9→C with. At the tRNA structure, the conjugates D3 cleavage points are indicated by arrows.

Conclusion

The described artificial ribonucleases represent a new family of tools to study RNA conformation under mild conditions maintaining the native structure of RNA. These conjugates represent simple mimics of RNase A: they contain structures capable of binding to RNA and residues catalyzing hydrolysis of phosphodiester bond. These simple molecules have an advantage as structural probes, as compared to RNase A. This cationic enzyme is known to affect RNA structure upon binding and in some cases it cleaves easily RNA at PypPu sequences in double-stranded regions of the molecules. In contrast to the enzyme, the mimics cannot unfold RNA. The developed artificial ribonucleases might be used for comparison of structures of mutant tRNAs and for probing structure of other RNA molecules. The efficiency of RNA cleavage by the developed compounds makes these molecules attractive prototype reactive groups for the design of second generation antisense oligonucleotide derivatives that could find novel applications for inactivation of RNAs in gene targeted therapeutics.

Acknowledgements

This work was supported by the Russian Foundation for Basic Research (projects Nos. 99-04-49538 and 96-15-97732), by INTAS (grant 96-1418), by Integration grant of Siberian Branch of Russian Academy of Science, and by Ministry of Education of Russian Federation (grant in the field of life sciences).

References

1. Konevetz, D.A., Beck, I.E., Beloglazova, N.G., Sulimenkov, V.S., Sil'nikov, V.N., Zenkova, M.A., Shishkin, G.V. and Vlassov, V.V. (1999). *Tetrahedron* 55, 503-514.
2. Zhdan N.S., Kuznetsova I.L., Vlassov A.V., Sil'nikov V.N., Zenkova M.A. and Vlassov V.V. (1999). *Nucleosides&Nucleotides* 18, 1491-1492.
3. Kolchanov, N., Vlassov, V. and Vlassova, I. (1996). *Prog. Nucl. Acids Res. & Mol. Biol*. 53, 131-196.
4. Beloglazova, N., Vlassov, A., Konevetz, D., Sil'nikov, V., Zenkova, M., Giege, R. and Vlassov, V. (1999). *Nucleotides&Nucleotsides* 18, 1463-1465.
5. Perret, V., Garcia J., Puglisi, Grosjean, H., Ebel J.-P., Florentz C. and Giege, R. (1990). Biochimie 72, 735-746

BINDING OF OLIGONUCLEOTIDES WITH GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE INTERFERES WITH CELLULAR FUNCTIONS OF THE ENZYME

Bryksin A.V., Laktionov P.P., Volod'ko N.V., Griffoni C[1]., Rykova E.Y., Spisni E[1]., Kraft R[2]., Tomasi V[1]., Vlassov V.V.

Institute of Bioorganic Chemisty SB RAS, Novosibirsk, Russia
[1]Institute of Cytomorphology, C.N.R., Bologna, Italy
[2]Max-Delbrueck-Centrum for Molecular Medicine (MDC), Berlin, Germany
e-mail: bryksin@niboch.nsc.ru
*Corresponding author

Keywords: glyceraldehyde-3-phosphate dehydrogenase, antisense oligonucleotides

Resume

Motivation:

Cellular uptake of oligodeoxynucleotides (ODN) is a key factors determining efficacy of this therapeutics capable of inhibiting expression of specific genes [1]. It was shown that cellular surface proteins are involved in oligonucleotides uptake [1]. A few putative ODN binding/transport proteins have been detected on the cell surface that could potentially mediate endocytosis of nucleic acids [1;2]. Investigation of interaction of ODNs with these proteins and intracellular nucleic acid binding proteins may shed light on the mechanisms of cellular uptake and nuclear accumulation of oligonucleotides.

Introduction

Antisense ODN 5'-pTACAGTAAATATCTAGGAATG (cPLA2p$(N)_{21}$) directed against cytosolic phospholipase A2 mRNA was shown to be efficiently taken up and enter into the nucleus of endothelial cells, human monocytes and HeLa cells whereas antisense ODN directed against PGH synthase-2 mRNA and control ODNs were found mainly in cytosole. Gel shift experiments show that cPLA2p$(N)_{21}$ binds to 37 kDa protein in nuclear extract of HeLa cells. TAAAT sequence was identified as the major binding motif for the 37 kDa nuclear protein [1].

In this study, 37 kDa ODN-binding protein was isolated from nuclear extract of HeLa cells and identified by peptide sequencing as glyceraldehyde-3-phosphate dehydrogenase (GAPDH). In experiments with this enzyme it was shown that the protein effectively binds anti-cPLA2 antisense ODN 5'-pCAGTAAATATCTAGGA (cPLA2p$(N)_{16}$). ODN-binding properties of GAPDH and effect of ODN on activity of this enzyme were studied. ODN-binding site on the GAPDH molecule was identified.

Materials and Methods

TAAAT containing cPLA2p$(N)_{16}$ sequence was synthesized on a ASM-102U DNA synthesizer (BioSet, Novosibirsk, Russia) by the phosphoramidite method and conjugated to alkylating reagent 4-[(N-2-chloroethyl-N-methyl)amino]benzylamine through its 5'-phosphate according to [2].

Nuclear extracts of HeLa cells was prepared and stored as described [5].

To estimate dissociation constant of ODN-protein complexes, nuclear extracts or commercial enzyme were incubated with 0.01-2.5 μM alkylating ODN conjugate [^{32}P]ClRpcPLAp$(N)_{16}$. The alkylating reaction products were separated by 10-20% SDS-PAGE. Concentration of alkylating ODN derivative was plotted as a function of the labeled protein band intensity.

Nuclear extract of HeLa cells was fractionated with ammonium sulphate (60-80%). The 60-80% pellet was dialysed against PBS and applied to affinity column with cPLAp$(N)_{16}$ covalently bound to Ultrogel A2.

Purified protein was sequenced by Edman degradation procedure.

In competition experiments, affinity modification of commercially obtained GAPDH (Sigma) with 1 μM [^{32}P]ClRpcPLAp$(N)_{16}$ was accomplished in the presence of 5x molar excess of ATP; GTP; AMP; GMP; NAD^+; NADH; glyceraldehyde-3-phosphate (G3P); cPLA2p$(N)_{16}$ or 15x weight excess of dsDNA; polyA; polyU; polyC; polyIC; heparin; dextran sulfate (DexS).

Hydrolysis of covalent ODN-protein adduct with hydroxylamine was performed as described elsewhere [1]. Glycolytic activity of free GAPDH and in presence of either 7,5 μM or 15 μM cPLA2p$(N)_{16}$ was assayed according to [2].

Results

To isolate the 37 kDa ODN-binding protein, HeLa cell nuclear extract was fractionated by stepwise addition of ammonium sulfate (AS) (Fig.1). The precipitates were dialyzed against PBS and incubated with 1μM [^{32}P]ClRpcPLA2p(N)$_{16}$. It was found that 37 kDa ODN-binding protein was precipitated with 80% AS. Proteins pelleted by stepwise addition of AS at 60-80% saturation were applied to affinity chromatography. Two protein bands with molecular masses of about 37 kDa were revealed in the 60-80% pellet; the upper band was found to interact with the affinity sorbent. The affinity purified protein was separated from minor contaminants by 10-20% SDS PAGE.

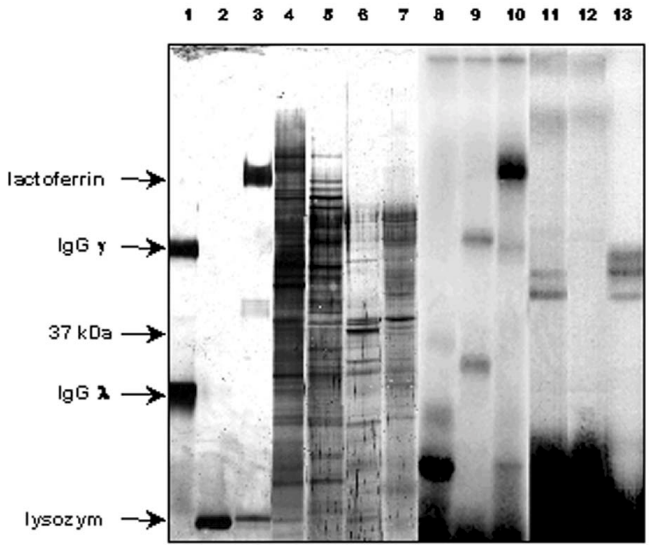


Figure 1. Isolation of TAAAT specific ODN-binding protein from nuclear extracts of HeLa cells.
lanes 1-7, colloidal silver staining of nitrocellulose blot;
lanes 8-13, autoradiography of the gel;
lanes 1, 2, 3 and 8, 9, 10 - molecular weight marker proteins;
lanes 4, 11 - nuclear extracts modified with [32P]ClRcPLA2p(N)16;
lanes 5, 12 - pellets AS 60% modified with [32P]ClRcPLA2p(N)16;
lanes 6, 13 - pellets AS 60-80% modified with [32P]ClRcPLA2p(N)16;
lane 7 - affinity purified 37 kDa protein.

9 peptides from tryptic digest of the 37 kDa protein have been sequenced. The peptides were found to correspond to 309-322, 200-214, 145-161, 227-233, 219-226, 5-12, 194-196, 248-258, 323-333 positions of aminoacid sequence of Human Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) Commercially obtained GAPDH modified with [^{32}P]ClRpcPLA2p(N)$_{16}$ showed the same electrophoretic mobility as the 37 kDa ODN-binding protein of HeLa nuclear extract.

Dissociation constant of ODN-protein complex was found to be 0.5 μM for the protein in the nuclear extract and 0.2 μM for commercial GAPDH. GAPDH displays high affinity to cPLA2p(N)$_{16}$ similar to that of other known proteins dealing with nucleic acids [2].

In order to investigate specificity of binding of cPLA2p(N)$_{16}$ with GAPDH, affinity modification of the protein was performed in the presence of various competitors (Fig.2). It was found that polyU, polyA, polyC, NAD$^+$ and Heparin are efficient competitors, while ATP, AMP, GTP, GMP, DexS, NADH, polyIC, dsDNA and G3P inhibited the affinity modification to a lesser extent.

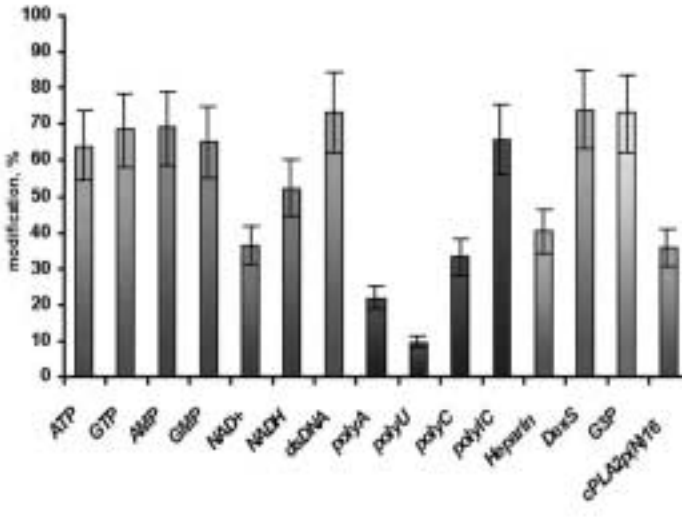


Figure 2. Reaction of GAPDH with 1□M [32P]ClRcPLA2p(N)16 in presence of 5x molar excess of ATP; GTP; AMP; GMP; NAD+; NADH; G3P ; cPLA2p(N)16 or 15x weight excess of dsDNA; polyA; polyU; polyC; polyIC; heparin; DexS were used.

Effect of ODN on dehydrogenase function of GAPDH was investigated. Gel shift experiments have shown that the enzyme forms complexes with ODN in the presence of 7,5(15) μM NAD+ and 7,5(15) μM ODN. Binding of ODNs to GAPDH led to negligible noncompetitive inhibition of dehydrogenase activity.

To localize the ODN-binding region on the GAPDH molecule, [^{32}P]ClRpcPLA2p(N)$_{16}$ was covalently bound to GAPDH and the enzyme was hydrolyzed with hydroxylamine. This treatment yielded two peptides with mol masses of 31 and 4,9 kDa. The label was associated with the 4,9 kDa peptide suggesting involvement of the peptide in the ODN-binding site of GAPDH.

Discussion

Classic glycolytyc protein GAPDH was found to display a number of activities including participation in membrane fusion, microtubule binding, phosphotransferase activity, nuclear RNA export, DNA replication and

DNA repair [2]. It is known that NH_2-domain of GAPDH (1-151) is responsible for NAD+ binding whereas COOH-domain (151-334) is responsible for binding of G3P (aminoacid residues 179, 231, 148, 195, 208, 209). The present investigation has shown that the dehydrogenase activity of the enzyme is not affected by cPLA2p$(N)_{16}$, whereas NAD+ prevents binding of oligonucleotide to the enzyme. NAD+ was shown to interfere with binding of RNA to GAPDH [2]. It should be noted that binding of GAPDH to NAD+ changes conformation of the enzyme [10]. Our data demonstrate that ODN-binding site of GAPDH is localized within aminoacid residues 286-334 that are not involved in G3P or NAD+ binding sites. Apparently, NAD+ and G3P interfere with oligonucleotide binding to GAPDH through inducing some conformational rearrangement of the enzyme.

Our data indicate that ODN-binding site of the enzyme can involve aminoacid residues from the region 286-334. It should be noted that GAPDH has nuclear localization (259-263) and nuclear export (202-208) signals that are outside of the ODN-binding sequence. These data enable us to suggest that binding of ODN with GAPDH does not interfere with nuclear shuttling of GAPDH, and this protein may be involved in intranuclear delivery of ODN.

Acknowledgements

The present work was supported by grants (40%) from MURST (Rome) to V.Tomasi and N.M Maraldi, Russian Foundation of Fundamental research N 99-04-49538 the grant from RCSTP "Gene directed biologically active compounds as anticancer and antiviruses drug" and INTAS grant no. 99-630.

References

1. Agrawal, S., Kandimalla, E.R. (2000) Antisense therapeutics: is it as simple as complementary base recognition? Mol. Med. Today, 6,72-81
2. Yakubov, L. A., Deeva, L. A., Zarytova, V. F., Ivanova, E. M., Ryte, A. S., Yurchenko, L. V., Vlassov, V. V. (1989) Mechanizm of oligonucleotide uptake by cells: involvment of specific receptors? Proc. Natl. Acad. Sci., 86, 6454-6459.
3. Bennett, R. M., Hefeneider, S. H., Bakke, A., Merritt, M., Smith, C. A., Mourich, D., Heinrich, M. C. (1988) The production and characterisation of murine monoclonal antibodies to a DNA receptor on human leucocytes. J. Immunol., 140, 2937-2942.
4. Hanss, B., Leal-Pinto, E., Bguggeman, L. A., Copeland, T. D., and Klotman, P. E. (1998) Identification and characterization of a cell membrane nucleic acid channel. Proc. Natl. Acad. Sci.,.95, 1921-1926.
5. Griffoni, C., Spisni, E., Orlandi, M., Santi, S., Riccio, M., Tomasi, V. A. (1999) 38 kDa nuclear protein is involved in the retention of an antisense oligonucleotide directed against cytosolic phospholipase A2. Nucleosides Nucleotides, 18, 1673-1676.
6. Blumenfeld, O. O., Rojrind, M., Gallop, P.M. (1965) Biochemistry, 4, 1780-1788.
7. Cori, G.T., Slein, M.W., and Cori, C.F. Crytalline (1948) D-glyceraldehyde-3-phosphate dehydrogenase from rabbit muscle. J.Biol.Chem., 173, 605-611.
8. Laktionov, P.P., Dazard, J.E., Vives, E., Rykova, E.Y., Piette, J., Vlassov, V.V. and Lebleu, B. (1999) Characterisation of membrane oligonucleotide-binding proteins and oligonucleotide uptake in keratinocytes. Nucleic Acids Res., 27, 2315-2324.
9. Sirover, M.A. (1999) New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase. Biochim. Biophys. Acta., 1432, 159-184.
10. Singh, R. and Green, M. R. (1993) Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase. Science, 259, 365-368.

THE MOLECULAR ANALYSIS OF MITOCHONDRIAL GENOME STRUCTURE OF WILD PERENNIAL CROP *ELYMUS SIBIRICUS*: EVOLUTIONARY CONSERVATION AND VARIATION OF COXI, COXIII, ATPA, RPS13 GENES

Tauson E.L., [1]Rogozin I.B., Verbitski D.S., *Konstantinov Yu.M., Arziev A.Sh., Bashalkhanov S.I.

Siberian Institute of Plant Physiology and Biochemistry SB RAS, Russia
[1]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: yukon@sifibr.irk.ru
*Corresponding author

Keywords: mitochondrial genes, wild crop, wheat, maize, evolution

Introduction

Plant mitochondrial genomes differ from those of other organisms. Their large size, the presence of repeated sequences involved in recombination processes, the existence of promiscuous chloroplast DNA sequences, the differences in genome organization even between closely related species and the presence of specific genes, especially the set of ribosomal protein genes, which are not found in nonplant mitochondrial genomes, are some of the characteristic features of plant mitochondrial genomes (Newton, 1988; Gualberto et al., 1990). For example, complete copies of genes encoding the ribosomal S12 protein are present in the mitochondrial genome of petunia, maize and wheat (Gualberto et al., 1988), whereas the portions of the analogous gene are deleted in the mitochondrial DNA of *Oenothera berteriana* (Schuster et al., 1990). Genes encoding the ribosomal S13 protein were found in tobacco, *Oenothera*, carrot, and maize (Bland et al., 1986; Schuster, Brennicke, 1987; Bonen, 1987), but are absent or not transcribed in the mitochondrial genome of such species as broad beans and pea (Bland et al., 1986; Schuster, Brennicke, 1987; Bonen, 1987). The gene encoding the ribosomal S14 gene is absent in the maize mitochondrial genome but is reported to occur in the genome of soybean, broad bean (Wahleithner, Wolstenholme, 1988), and *Oenothera berteriana* (Schuster et al., 1990) mitochondria. The biological consequence of localization of mitochondrial ribosomal protein genes in different genetic systems of the plant cell still remains obscure.

Until recently, the mitochondrial genome of wild perennial crops has not been an object of intensive molecular biological studies as opposed to cultivated plant species. Yet, the functional organization of the mitochondrial genome in species of these crops is of interest because, in contrast to cultivated plants, it failed to undergo a rigorous selection for development of certain economically valuable traits.

Present study is intended to analyze the primary structure of some mitochondrial genes of wild crop *Elymus sibiricus* L.: full nucleotide sequence of rps13 gene and fragments of cytochrome oxidase subunit I and III (coxI, coxIII), subunit A of ATP synthase. Phylogenetic analysis of nucleotide sequences of rps13, coxI, coxIII genes in plant species was carried out.

Results and Discussion

The genomic nucleotide sequence of the perennial wild crop *E. sibiricus* rps13 gene shows very strong homology to the S13 ribosomal protein gene of the wheat (Bonen, 1987). The genomic nucleotide sequences of *E. sibiricus* and wheat rps13 genes are 99.7% identical. The only difference was found in position 45 of the sequence (adenine in the wheat instead of cytosine). The comparison of rps13 nucleotide sequence of *E. sibiricus* with genome sequence of the corresponding genes of *Daucus carota* and *Oenothera berteriana* (Wissinger et al., 1990) shows its comparatively high homology (93,9 and 94,9 % respectively). Besides, the following editing of rps13 transcripts in *Daucus carota* mitochondria results in decrease of similarity to the *nt* sequence of *E. sibiricus* gene to 93.0% because of the appearance of additional differences at positions 26, 56 and 101 (C to T conversion). The editing of the mitochondrial *rps* transcripts in *Oenothera berteriana* also results in decrease of it's homology to *E. sibiricus* sequence to 93.8%, due to C to T conversions at positions 26 and 287 (Wissinger et al., 1990).

As a whole, the results of the comparison of the nucleotide sequences are in total agreement with a taxonomic position of the species compared. It should be noted that the most differences between amino acid sequences of S13 of *Daucus carota* and *Oenothera berteriana* occur in the N-terminal half of the polypeptide, while in *E. sibiricus* S13 polypeptide these differences are distributed through the length.

The deduced 116 *aa* sequence of *E. sibiricus* was compared with those of the wheat (Bonen, 1987), maize, tobacco (Bland et al., 1986) and carrot (Wissinger et al., 1990). The lengths of the sequences were similar

varying from 116 to 129 *aa* (Fig. 2A). The *E. sibiricus* sequence has a complete homology to that of wheat and slightly lower, though still rather high (98%) to the maize sequences. Besides, the comparison demonstrates the existence of a rather high similarity of the rps13 protein of *E. sibiricus* to tobacco (91%) and carrot (89%) sequences. As a result of transcript editing in carrot mitochondria, the homology of the final polypeptide product with that of *E. sibiricus* decreases to 86% by amino acid substitutions in positions 9 (S to L), 19 (S to L) and 34 (R to C) of the amino acid sequence (Wissinger et al., 1990). The degree of homology of the sequence of *E. sibiricus* to that of a moss specimen *Marchantia polymorpha* and bacterium *Escherichia coli* was 62 and 44%, respectively.

Further analysis of the deduced *aa* sequence of the rps13 protein revealed unknown hydrophobic amino acid motif -L-X10-L-X10-M-X10-L-X10-L- at *aa* positions 37-81 (fig. 2, B). The motif characteristically differs from the known leucine motifs in proteins of different origin including mitochondrial one (Konstantinov, Moller, 1994) by the disruption of leucine regularity of the motif by incorporation of another hydrophobic amino acid, methionine, in the position 59 of the amino acid sequence.

Protein secondary structure predicted by the Garnier method (Garnier et al., 1978) shows that the region including the motif has an alpha-helical conformation. Since leucines (including one methionine) 11 *aa* appear on the same side of an alpha-helical wheel, this suggests that new hydrophobic amino acid motif is perhaps involved in ribosomal protein-protein interactions through helix-helix association by analogy with the leucine zipper motif.

It should be mentioned that similar motif is also found in other ribosomal proteins of mitochondria and chloroplasts. A special search with the use of EMBL database revealed that the motifs formed with involvement of some specimens of hydrophobic amino acid with a long side chain (leucine, methionine, isoleucine) are rather typical for the *aa* sequences of ribosomal proteins in mitochondria and chloroplasts.

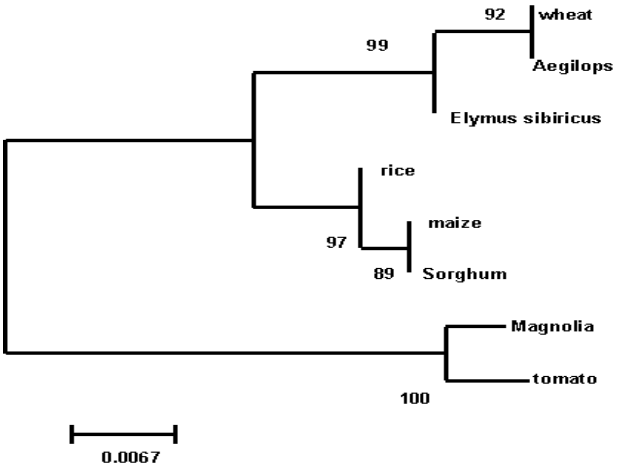


Figure 1 shows the phylogenetic tree of cox I gene. It may be seen that Elymus sibiricus forms monophiletic group together with Triticum aestivum and Aegilops columnaris. This is in a good agreement with the results of bootstrap test (P=0.99) for this br

The data on distribution of the motif in the ribosomal proteins of mitochondria and chloroplasts permit us to assume that this *aa* motif provides folding and binding of particular proteins in the ribosomal supramolecular complex.

Phylogenetic analysis of nucleotide sequences was performed by using the computer programs MEGA 1.03 (Kumar, Tamura, and Nei 1993) and PAUP* 4.0 (Swofford 1998). The neighbor-joining and maximum parsimony methods were used to reconstruct phylogenetic trees from nucleotide sequence data (amino acid sequences were not informative due to low divergence of sequences studied). The bootstrap test was conducted with 1000 replications. The mitochondrial coxI, coxIII, atpA, rps13 genes sequences of higher plants were taken from GenBank (section PLN).

Figure 1 shows the phylogenetic tree of cox I gene. It may be seen that *Elymus sibiricus* forms monophiletic group together with *Triticum aestivum* and *Aegilops columnaris*. This is in a good agreement with the results of bootstrap test (P=0.99) for this branch of phylogenetic tree. The trees constructed by using the neighbor-joining and maximum parsimony methods have an identical topology which points on adequacy of phylogeny constructed.

It is interesting that topology of tree for coxI coincides generally with the topology of tree for rps13 gene (Fig. 2). It may be seen that Elymus sibiricus forms group together with Triticum aestivum with a good statistics (P=0.93).

It is difficult to carry out phylogenetic analysis of mitochondrial genes in plants because of extremely low rate of mutation fixation during their evolution. But even

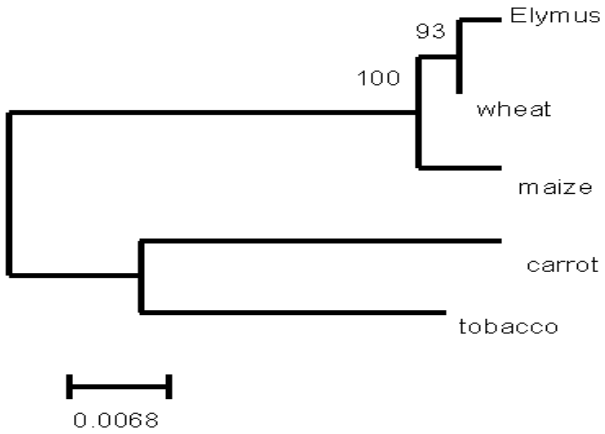


Fig. 2. Phylogenetic tree for rps13 gene. The numbers for the interior branches refer to the bootstrap values. The Magnolia and tomato sequences used as outgroup sequences under determination of phylogenetic tree root.

in the case of low level of sequences divergence the phylogenetic trees obtained by us were rather stable. We suggest that on the basis of mitochondrial genes characteristics genus *Elymus* and *Triticum* are evolutionary close to each other in comparison with genus *Zea*. At present it is rather difficult to evaluate quantitatively this closeness. We showed that proportion of nucleotide sites at which two nucleotide sequences are different for *Elymus* and *Triticum* is sufficiently lower (3-4-fold) that for *Elymus* and *Zea* (Tab. 1).

Table 1. The proportion of nucleotide sites at which two nucleotide sequences are different for coxI (right upper corner) and rps13 (left low corner).

	Elymus	***Triticum***	***Zea***
Elymus		0.006	0.022
Triticum	0.003		0.028
Zea	0.012	0.009	

The study is supported by the grant INTAS No. 98-0522.

References

1. Bland, M.M., Levings, III C.S., Martzinger, D.F. 1986. The tobacco mitochondrial ATPase subunit 9 gene is closely linked to an open reading frame for a ribosomal protein. Mol. gen. Genet. **204:** 8-16.
2. Bonen, L. 1987. The mitochondrial S13 ribosomal protein is silent in wheat embryos and seedlings. Nucleic Acids Res. **15**: 10393-10404.
3. Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **38**: 16-24.
4. Garnier, J., Osguthorpe, D.J., Robson, B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**: 97-120.
5. Gualberto, J.M., Wintz, H., Weil, J.H., Grienenberger, J.M. 1988. The genes coding for subunit 3 of the NADH dehydrogenease and for ribosomal proreins S12 are present in the wheat and maize mitochondrial genomes and are co-transcribed. Mol. Gen. Genet. **215**: 118-127.
6. Gualberto, J.M., Weil, J.H., Grienenberger, J.-M. 1990. Editing of the wheat coxIII transcript: evidence for twelve C to U and U to C conversion and for sequence similarities around editing sites.Nucleic Acids Res. **18**: 3771-3776.
7. Konstantinov, Yu.M., Moller, I.M. 1994. A leucine motif in the amino acid sequence of subunit 9 of the mitochondrial ATPase, and other hydrophobic membrane proteins, that is highly conserved by editing. FEBS Lett. **354:** 245-247.
8. Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: Molecular Evolutionary Genetics Analysis, Version 1.02. The Pennsylvania State University, University Park, PA.
9. Newton, K. 1988. Plant mitochondrial genomes: organization, expression and variation. Ann. Rev. Plant Physiol. Plant Mol. Biol. **39**: 503-532.
10. Quetier, F., Lejeune, B., Delorme, S., Falconet, D. 1985a. Encyclopedia of Plant Physiol. - Berlin-Heidelberg. **18**: 25-36.
11. Quetier, F., Lejeune, B., Delorme, S., Falconet, D. 1985b. Molecular Form and Function of Plant genome. - N.Y. **83**:413-420.
12. Schuster, W., Brennicke, A. 1987. Plastid DNA in the mitochondrial genome Oenothera. Mol. Gen. Genet.. **210**: 44-51.
13. Schuster, W., Unseld, M., Wissinger, B., Brennicke, A. 1990. Ribosomal protein S14 transcripts are edited in Oenothera mitochondria. Nucleic Acids Res. **18**: 229-233.
14. Swofford, D.L. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Suderland, Mass.
15. Wahleithner, J.A., Wohstenholme, D.R. 1988. Ribosomal protein S14 genes in broad bean mitochondria. Nucleic Acids Res. **14**: 6897-6913.
16. Wissinger B, Schuster W, Brennicke A. 1990 Species-specific RNA editing patterns in the mitochondrial rps13 transcripts of Oenothera and Daucus. Mol Gen Genet., **224**:389-95.

AGE-ASSOCIATED CHANGES IN 8-OXOGUANINE- AND HYPOXANTHINE-DNA-GLYCOSYLASE ACTIVITIES OF CELL'S EXTRACTS OF WISTAR AND OF RATS WITH INHERITED OVERGENERATION OF FREE RADICALS

*[1]Ishchenko A.A., [2]Sinitsyna O.I., [2]Vasunina E.A., [2]Krysanova J.S., *[1]Nevinsky G.A.*

[1]Institute of Bioorganic Chemistry SB RAS, Novosibirsk, Russia
[2]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: nevinsky@niboch.nsc.ru
*Corresponding author

Keywords: 8-oxoguanine, nuclear and mitochondrial extracts, 8-oxoguanine-DNA-glycosylase, substrate specificity, rats, ageing, OXYS

Resume

Motivation & Results:

Oxidative stress is one of the important factors which leads to DNA lesions resulting in mutations, cancer and ageing. A variety of mutagenic and cytotoxic DNA lesions are formed by oxidative DNA damage, one of the most abundant beings 8-oxoguanine (oxoG). A technique for determination of the level of 8-oxoG- and hypoxanthine-DNA-glycosylase activities of nuclei and mitochondrial lysates of hepatocytes of Wistar and of rats with inherited overgeneration of free radicals (OXYS strain) was developed. Age-associated changes in the glycosylase activities of cell's extracts of Wistar and OXYS rats were detected. The pronounced difference in specificity of mitochondrial and nuclear glycosylase activity was revealed.

Introduction

Free oxygen radicals arising abundantly under oxidative stress: superoxide radicals and hydroxyl radicals, induce premutational modifications of DNA bases leading to formation of 8-oxo-deoxyguanosine (8-oxo-dGua), 8-oxo-deoxyadenosine (8-oxo-dAde), thymine diols, FapyGua, and FapyAde [1]. It was shown that inherited overgeneration of free radicals in animals is accompanied by a number of morbid conditions resembling human degenerative diseases of ageing. Some of those (cataracts, cardiomyopathy, emphysema, carcinogenesis and others) as well as short life-span were determined in OXYS rats [2]. Thus, the OXYS strain could be validated as a model to study human degenerative diseases. In this study age-associated changes in the glycosylase activities of cell's extracts of OXYS compare to Wistar rats were demonstrated.

Methods

Oligodeoxyribonucleotides (ON) containing oxoG (**G***) or hypoxanthine (**H***) residue in different positions were used for measuring glycosylase/AP-lyase activity in crude nuclear or mitochondrial extracts. Cytoplasmic extracts have shown no marked activity.

The ONs marked as OG**N** or Hyp**N** , where **N** is the distance from the ON 5'-end to the oxoG or hypoxanthine residue accordingly:

OG6, CTCTC**G***CCTTCCTCCTTTCCTCT;

G6, CTCTCGCCTTCCTCCTTTCCTCT;

OG11, CTCTCCCTTC**G***CTCCTTTCCTCT;

Hyp6, CTCTC**H***CCTTCCTCCTTTCCTCT.

23-mers containing modified bases were [5'-^{32}P]-labelled and annealed to the complementary sequences possessing C positioned opposite the oxoG base and T opposite the Hyp.

Standard reaction mixtures (30 μl) contained 20 mM HEPES-NaOH, pH 7.5, 80 mM NaCl, 5 mM EDTA, 3 mM DTT, 9% glycerol, 50 nM of OG6 (or 100 nM of Hyp6), and 0,075 mg/ml mitohondrial or nulear protein, were incubated at 25^{0}C for 1 h. The reaction products were resolved by PAGE in 20% gel containing 7 M urea.

Results and discussion

Fig. 1 demonstrates a typical gel of OG6 and Hyp6 degradation by the nuclear or by the mitochondrial extracts (of the rat's hepatocytes). The cleavage products of OG6 after incubation with nuclear extracts under these experimental conditions co-migrate with product of OG6 incubation with purified mice's mOGG1 protein (β-elimination product).

The products migrating as a doublet in PAGE with TBE buffer are typical for β-elimination of ONs [3]. In the case of the mitochondrial extracts one additional [5'-^{32}P]ON was observed which co-migrate with the product of OG6 digestion by purified *E. coli* Fpg protein (β,δ-elimination product) (Fig. 1). No significant activity on single stranded OG11 or unmodified G6 duplexes was revealed in either nuclear or mitochondrial extracts (data not shown).

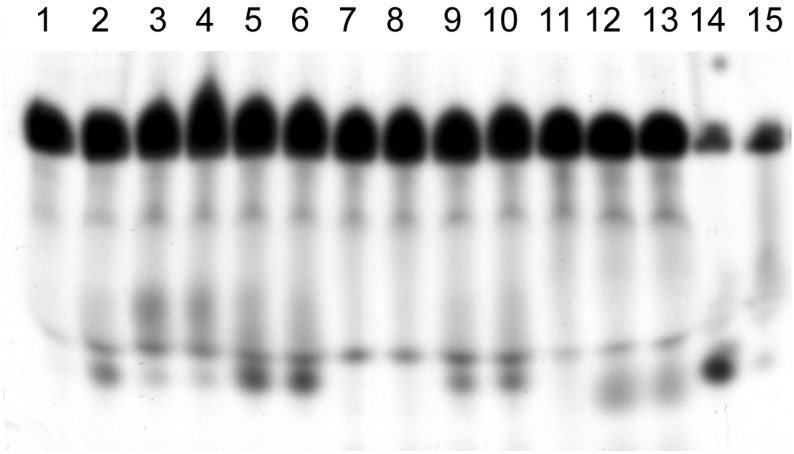


Figure 1. Urea-PAGE analysis of the dsOG6 (lines 1-10) and dsHyp6 (lines 11-13) hydrolysis b prepared from rats livers (radioautograph). Lines 1 and 11 - control incubation of dsOG6 and H Lines 2, 14 and 15 – controls as in line 1 but with additional treatment with NaOH, Fpg and mO 4 - hydrolysis of dsOG6 by Wistar and OXYS nuclear extracts respectively using standard con 6 – as 3 and 4 but with additional treatment with NaOH respectively. Lines 7 and 8 - hydrolysis mitochondrial extracts using standard conditions respectively. Lines 9 and 10 – as 7 and 8 but respectively. Lines 12 and 13 - hydrolysis of dsHyp6 by Wistar and OXYS mitochondrial extrac conditions.

There were defined conditions for incisions at the AP-sites formed after the oxoG removal in oligonucleotides treated with the nuclear extracts since the nuclear mOGG1 glycosylase has AP-lyase activity weaker then its 8-oxoG-glycosylase activity. The ON containing an AP-site at sixth position from 5'-end was obtained by Hyp6 incubation with ANPG protein. Treatment of the ON with 30 mM NaOH and Fpg protein shows that all AP-sites hydrolysed by 30 mM NaOH after 2 min at 90^{0}C. Moreover, two types of the AP-site hydrolysis were compared: treating with 0.1 M putrescine and with 30 mM NaOH. For the case with putrescine (and for piperidine as well) the greater cleavage of control ON at oxoG site than with NaOH was detected (data not shown).

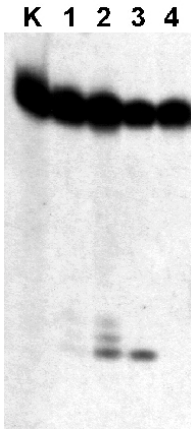


Figure 2. Urea-PAGE analysis of ds [5'-32P]OG11 hydrolysis by nuclear (1, 2) and mitochondrial extracts (3, 4) prepared from Wistar liver (radioautograph): 1 and 3 - after 15 min of the incubation; 2, and 4 - after 1 h of the incubation. K, the oligonucleotide incubated for 1 h.

Subsequent NaOH treatment of OG6, as well as Hyp6, after incubation with mitochondrial, contrary to incubation with nuclear extracts, does not result in additional labelled ON products formation in comparison with control (Fig. 1). These results indicate that under used conditions mitochondrial AP endonuclease activity is not a rate-limiting step.

Probably, the DNA-repair rate of the nuclear extracts might be as well defined due to the rate of dsOG6(11) cleavage without alkaline treatment, but as a result of joint activity of nuclear glycosylase and AP-endonuclease/lyase.

The on-time non-linear accumulation of dsOG11 cleavage product and its further disappearing after long incubation with the mitochondrial extracts was observed (Fig. 2). And it was not observed while using dsOG6. Therefore, it can be explained by a high post-repair exonuclease activity of the mitochondrial extracts. Such activity was actual for the OG11 duplexes only, since apart from those OG6 they were not melted after removal of the oxoG.

The conditions were defined for the cleavage of dsOG6 by the nuclear and mitochondrial extracts: concentration of dsOG6, the extracts protein; time of incubation. Selected were values at which all corresponding dependencies fell into linear region.

Fig. 3A demonstrate that 8-oxoguanine-DNA-glycosylase activity of OXYS in comparison with Wistar mitochondrial extracts was 1,25 and 1,36 times more higher at 3 and 12 month rats respectively and there wasn't significant differences at 6 and 15 month of age. The values show a slight decrease in OXYS incision activity (~23%) at age 6 months, relative to 3 months. There is a significant increase (~100%) in OXYS and Wistar mitochondrial incision activities at 15 months, relative to 3 months. Similar age-associated changes in hypoxanthine-DNA-glycosylase activity of OXYS and Wistar mitochondrial extracts were revealed (Fig. 3B). The

level of hypoxanthine-DNA-glycosylase activity was lightly but not significantly higher than 8-oxoguanine-DNA-glycosylase activity in the mitochondrial extracts. Although we measured cleavage of an 8-oxodG-containing substrate by whole mitochondrial extracts, it is very likely that this activity reflects the activity of mtODE, which was previously isolated from Wistar liver mitochondria [4].

8-oxoguanine-DNA-glycosylase (rOGG1) activity in crude nuclear extracts was measured as in the case of the mitochondrial extracts but with additional NaOH treatment after incubation as described above. The results obtained did not revealed differences between OXYS and Wistar glycosylase activity, but the age-associated changes in the activity were identical to those of mitochondrial extracts (Fig. 3C). The average 8-oxoguanine-DNA-glycosylase activity in nuclear was ~2 times higher than in mitochondrial extracts. There is no data for nuclear hypoxanthine-DNA-glycosylase activity. It was evidenced that single freezing of the extracts at -70^{0}C did not influence on their 8-oxoguanine-DNA-glycosylase activity, but can dramatically decrease nuclear hypoxanthine-DNA-glycosylase activity.

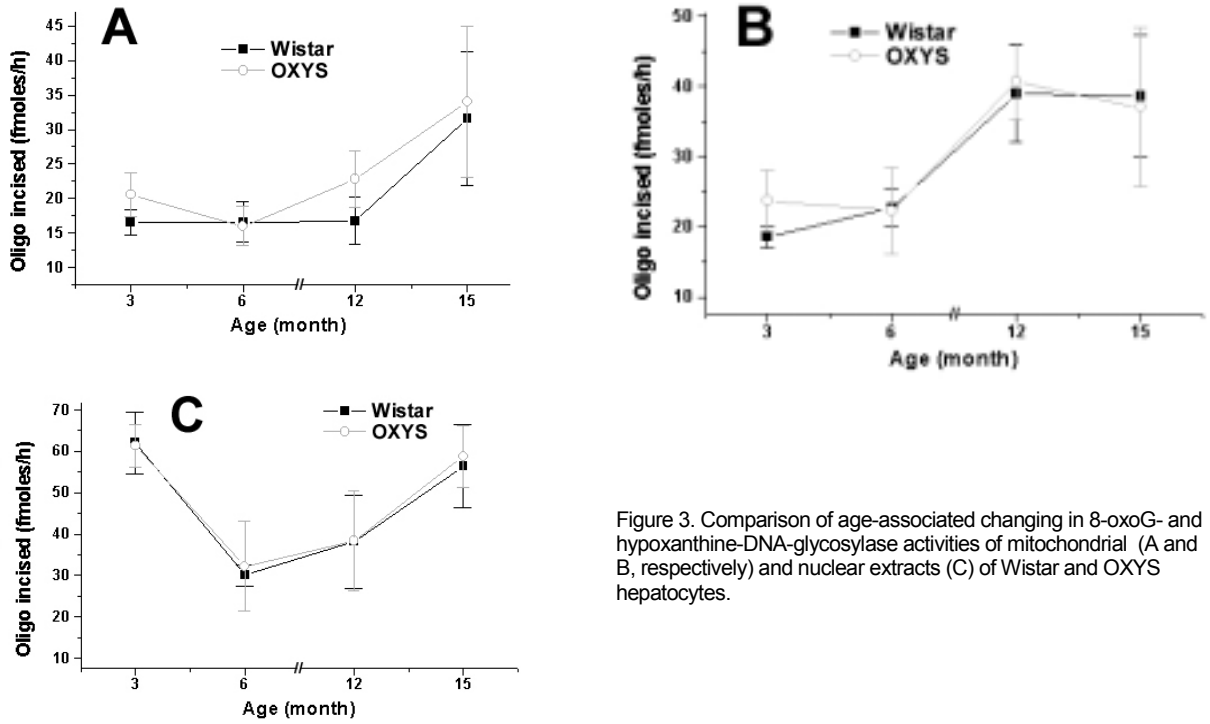


Figure 3. Comparison of age-associated changing in 8-oxoG- and hypoxanthine-DNA-glycosylase activities of mitochondrial (A and B, respectively) and nuclear extracts (C) of Wistar and OXYS hepatocytes.

Thus, in contrary to notion that the DNA repair capacity declines with age, we find a significant increase in OXYS and Wistar mitochondrial and nuclear incision activity with age upwards of 6 month in liver of the rats. The same results were obtained from measuring activity of catalase and from estimation of relative amount of carbonyl groups in protein of the cell's extracts as biochemical markers, which could be associated with the oxidative stress in these animals.

Acknowledgements

The research was made possible in part by a grant from INTAS No 99-1645, from The Welcome Trust (England), from the Russian Foundation of Basic Research and a grant from the Siberian Division of Russian Academy of Sciences.

References

1. Laval, J. (1996) Role of DNA repair enzymes in the cellular resistance to oxidative stress, Pathol. Biol. (Paris), 1, 14-24.
2. Salganik R.I., Solovyova N.A., Dikalov S.I., Grishaeva O.N., Semenova L.A. and Popovsky A.V. (1994) Inherited enhancement of hydroxyl radical generation and lipid peroxidation in the S strain rats results in DNA rearrangements degenerative diseases and premature aging, *Biochem. Biophys. Res. Comm.*, 199, 726-733.
3. Bailly, V., Verly, W.G., (1987) Escherichia coli endonuclease III is not an endonuclease but a beta-elimination catalyst. *Biochem. J.,* **242**, 565-572.
4. Croteau, D.L., ap Rhys, C.M., Hudson, E.K., Dianov, G.L., Hansford, R.G., Bohr, V.A. (1997) An oxidative damage-specific endonuclease from rat liver mitochondria., *J. Biol. Chem.,* **272**, 27338-44.

BISPYRENYLATED OLIGONUCLEOTIDE PROBES DISPLAYING HYBRIDIZAION INDUCED EXCIMER FLUORESCENCE

Kostenko E.V., Dobrikov M.I., Komarova N.I., Vlassov V.V., *Zenkova M.A.

Institute of Bioorganic Chemistry SB RAS, Novosibirsk, Russia
e-mail: Marzen@niboch.nsc.ru
*Corresponding author

Keywords: fluorescent probes, excimer fluorescence, oligonucleotides, RNA structure, hybridization

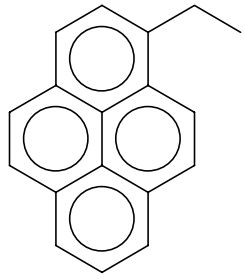
Resume

Design of fluorophore labeled oligonucleotides has been the subject of extensive investigations, since this type of oligonucleotide derivatives can be used for solution based assay of nucleic acids sequences and structures [1-4]. Pyrene excimer emission provides a powerful tool for studies of interactions involving RNA an DNA [5]. Several attempts to use pyrene excimer fluorescence as a probe for duplex formation between pyrene-labeled oligonucleotides have been failed or complicated by multy-step synthesis of the oligonucleotide conjugates [6]. We present here simple and reliable synthesis of 5'-bispyrenylated oligonucleotides and results of studies on hybridization of the labeled oligonucleotides with complementary DNA and RNA targets.

Results and discussion

5'-bis-pyrenylated oligonucleotide derivatives were synthesized using one step procedure from 1-pyrenylmethylamine hydrochloride through activation of 5'- phosphate of oligonucleotide by dipyridyldisulfide, triphenylphosphine and dimethylaminopyridine in dry DMSO. Very short spacer used in the conjugates prevents intercalation of the pyrene residue between base pairs in duplex. The yield of bis-pyrenylated oligonucleotide was 82-98%. Mono-fluorophore labeled oligonucleotide was formed as by-product with yield of 2-8%. Characterization of the obtained ON-I^{bis} and ON-II^{bis} oligonucleotide conjugates using polyacrylamide gel electrophoresis, HPLC, UV- and fluorescence spectroscopy and LC-MC confirmed the presence of two pyrenyl-1-methylamine residues at the 5'- phosphate of the oligonucleotides. Oligonucleotides and corresponding complementary DNA sequences are shown in Table 1.

Table 1. Sequences of bis-pyrenylated oligonucleotides and target DNAs.

 CH_2-NH_2 pyrenyl-1-methylamine (®)	5' d(CCAAGGAGCGCGAGGTCG) Tg-I 3' d(GTTCCTCGCGCTCCAG)-®$_2$ ON-I^{bis} 5' d(TGAATCTGGAGGAAGAGA) Tg-II 3' d(CTTAGACCTCCTTCTC)-®$_2$ ON-II^{bis} 5' d(GTCCAGCCCCATGGA)-®$_2$ ON-III^{bis} 5' d(CATCCATGGGGCTCCACTT) Tg-IV

The bis-pyrenylated oligonucleotides exhibit fluorescence typical of pyrene monomer emission (λ_{max}=382 nm) and additional excimer fluorescence band (λ_{max}=476 nm) (Fig. 1). Excimer fluorescence is extremely sensitive to environment: duplex formation between bis-pyrenylated oligonucleotide conjugates and complementary DNA causes 8-10-fold increase in excimer fluorescence intensity (Fig. 1). Changes in excimer fluorescence upon binding with target sequences depends directly on the amount of the target DNA in solution in the range of concentrations from 10^{-6} M to 10^{-9} M. Sensitivity of the assay is limited by the value of the equilibrium association constant of the duplex which was measured to be $2x10^7$ M^{-1} and $1.7x10^7$ M^{-1} for ON-I^{bis} and ON-II^{bis}, respectively. Addition of non-complementary DNA (Tg-IV, Table 1) to bis-pyrenylated oligonucleotides ON-I^{bis} and ON-II^{bis} resulted in neither hypochromicity, nor a changes in fluorescence spectra of the conjugates.

Thus the increase in excimer fluorescence intensity observed upon interaction with complementary DNA is the consequence of duplex formation. The effect of pyrene label on thermodynamic parameters of duplex formation between oligonucleotide and complementary DNA was studied by optical melting at 280 nm. Determined T_m^{s} were similar for duplexes formed by the conjugate and by oligonucleotide, indicating, that the modification does not alter interaction of the oligonucleotides. Identification of optimal sites for oligonucleotides hybridization is one

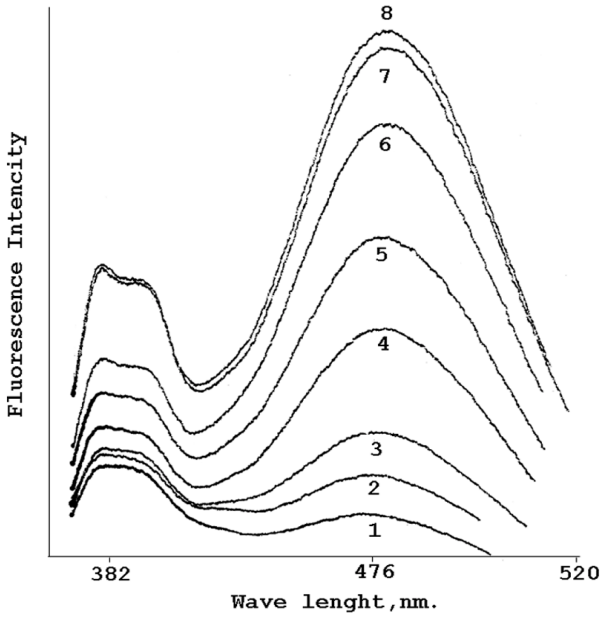


Figure 1 Fluorescence monitoring titration of 5'-bis-pyrenylated oligonucleotide ON-Ibis with complementary target DNA in 50 mM HEPES, pH 7.5, containing 5 mM MgCl2, 0.2 M KCl, 0.1 mM EDTA at 37 oC. □ex was 343 nm. 1 – fluorescence emission spectrum of ON-Ibis (5x10-7 M), 2 - 8, oligonucleotide ON-Ibis in the presence of complementary DNA, 10-7 M, 2x10-7 M, 5x10-7 M, 7x10-7 M, 10-6 M, 2x10-6 and 3x10-6 M, respectively.

of the major problems of the antisense approach. Fluorescence oligonucleotide conjugates are useful tools for investigation of antisense oligonucleotides interaction with their targets. We used the bis-pyrenylated oligonucleotides for identification of accessible regions in the *mdr1* mRNA and for investigation of kinetics of antisense oligonucleotides binding.

5'-region of human *mdr1* gene mRNA, which is responsible for resistance of tumor cells to chemotherapy, was used as the target. The investigated oligonucleotides were targeted to different parts of the *mdr1* mRNA Binding of bispyrenylated oligonucleotides to the *in vitro* transcript of the *mdr I* mRNA was monitored using fluorometric-titration assay. Solution of 5'-bispyrenylated oligonucleotide was titrated with increasing concentrations of mRNA.. The oligonucleotide conjugates displayed different fluorescence changes upon binding to different sites (Fig.2, Fig.3).

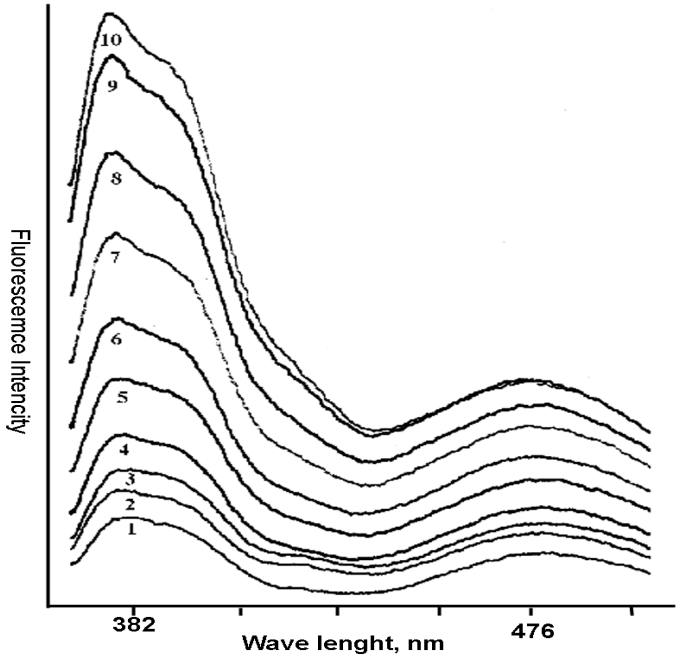


Figure 2. Fluorescence emission spectra of the bis-pyrenylated deoxyoligonucleotide ON-IIbis upon binding with in vitro transcript of mdr I mRNA in 50 mM HEPES, pH 7.0; containing 200 mM KCl, 5 mM MgCl2 and 0,1 mM EDTA at 370C. 1 - free bispyrenilated oligonucleotide ON-IIbis 5x10-7 M. 2 - 10 - ON-IIbis in the presence of mdr I mRNA; for mRNA concentrations 1x10-7 M, 3x10-7 M, 5x10-7 M, 7x10-7 M, 1x10-6 M, 2x10-6 M, 2,5x10-6 M, 3x10-6 M, 4x10-6 M, for spectra 2 - 10, respectively.

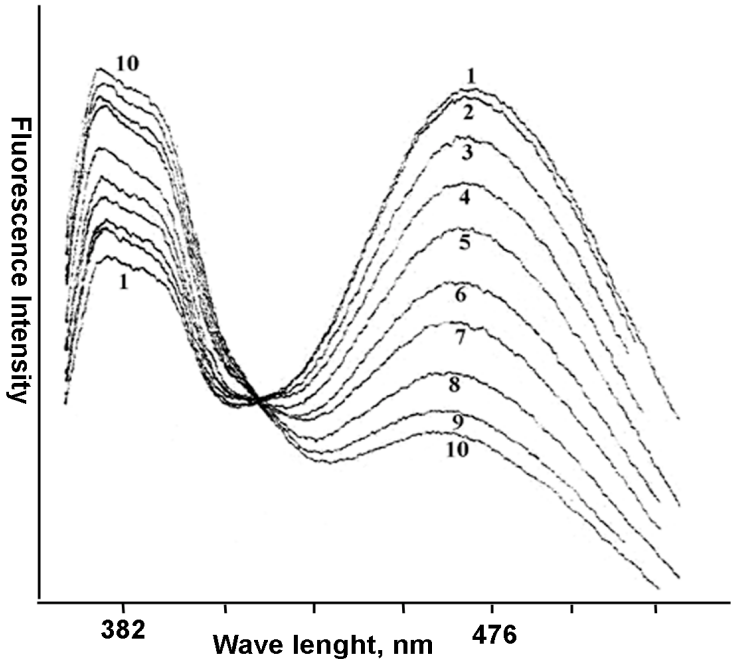


Figure 3. Fluorescence emission spectra of bis-pyrenylated deoxyoligonucleotide ON-IIIbis in the presense of in vitro transcript of mdr I mRNA in 50 mM HEPES, pH 7.0; containing 200 mM KCl, 5 mM MgCl2 and 0.1 mM EDTA at 370C. 1 - free pyrene-labeled oligonucleotide ON-IIIbis (5x10-7 M). 2 - 10 - ON-IIIbis in the presence of mdr I mRNA. mRNA concentrations were 1x10-7 M, 3x10-7 M, 5x10-7 M, 7x10-7 M, 1x10-6 M, 2x10-6 M, 2.5x10-6 M, 3x10-6 M, 4x10-6 M, for spectra 2 - 10, respectively.

Conjugate ON-II^{bis} which exhibited 10-fold increase in excimer fluorescence upon binding with complementary DNA (Fig. 1) displayed 6-fold enhancement in monomer fluorescence and 2.4 folds increase in excimer fluorescence upon binding to RNA (Fig. 2). Different fluorescence changes were displayed by oligonucleotide ON-III^{bis} (Table1, Fig. 3): monomer fluorescence increased (1.5 folds) whereas excimer fluorescence quenched (2.2 folds) upon binding to RNA. Thermodynamic and kinetic parameters of the oligonucleotides binding to RNA were determined using differences in fluorescence of the free and RNA bound oligonucleotide conjugates. Association constants for bis-pyrenylated oligonucleotide ON-I^{bis}, ON-II^{bis}, ON-III^{bis} binding to RNA were $1x10^6$ M^{-1}, $0.77x10^6$ M^{-1}, $0.2x10^6$ M^{-1}, respectively.

The results evidence, that excimer fluorescence of the developed oligonucleotides is extremely sensitive for microenvironment: conjugates show the excimer fluorescence (about 8-10 folds) upon binding to complementary DNA. In the case of binding to RNA monomer fluorescence increase was observed. Sensitivity of excimer fluorescence of bis-pyrenylated oligonucleotides on the structure of the binding site in RNA may provide a possibility to use these probes for investigation of high order structure of RNA.

Acknowledgments

We thank Dr. M. Manoharan for helpful discussion and Dr. H. Gaus for LC-MC analysis of bis-pyrenylated conjugates. This work was supported by research grant from INTAS (INTAS-RFBR 95-0653), from Russian Foundation of Basic Research (RFBR 99-04-49559, Interdisciplinary grant from Siberian Branch of Russian Academy of Science.

References

1. Paris, P., Langenhan J.M., Kool, E.T. (1998) Probing DNA sequences in solution with a monomer-excimer fluorescence color change. Nucleic Acids Res., 26, 3789-3793.
2. Yamana, K. et al. (1997) Tetrahedron Lett*., 53*, 4265-4270;
3. Telser, J. et al. (1989) J. Am. Chem. Soc. 111, 6966-6976.
4. Kierzek, R., Li Yi, Turner, D., Bevilacqua, P.C. (1993) 5'-Amino Pyrene Provides a Sensitive, Nonpertubing Fluorescent Probe of RNA Secondary and Tertiary Structure Formation. J. Am. Chem. Soc.115, 4985-4992.
5. Turro, N. J. *Modern molecular photochemistry*, Benjamin/Cammings: MelnoPark, CA, 1978, 141-143.
6. Tong, G., Lawlor, J.M., Tregear G. W., Halambardis, J. (1995) Oligonucleotide-polyamibe hybride molecules containg multiple pyrene residues exhibit significant excimer fluorescence. J. Am. Chem. Soc, 117, 12151-12158.

CHEMICAL CLEAVAGE OF PLASMID DNA BY ARENES IN THE PRESENCE OF CU(II) IONS

Koval O.A., Oleinikova S.B.[1], Chernolovskaya E.L., Litvak V.V., Vlassov V.V.

Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Institute of Bioorganic Chemistry SB RAS, Novosibirsk, Russia
e-mail: o.koval@niboch.nsc.ru
*Corresponding author

Keywords: artifical nucleases, DNA cleavage, Cu(II) ions

Resume

Motivaton:

Design of chemical compounds capable of sequence specific cleaving DNA can result in development of therapeutics exerting their action at the level of genome functioning. Targeting of specific sequences in DNA can be achieved by using oligonucleotides and peptides. A number of compounds capable of cleaving DNA have been conjugated to oligonucleotides [Knorre, 1994]. However the designed conjugates demonstrated poor activity in physiological condition. The major problem that remains to be solved is the design of small molecules capable of efficiently cleaving DNA under physiological conditions.

Results:

In the presence of Cu(II) ions, plasmid DNA is cleaved under physiological condition by different arenes at low concentrations. Supercoiled plasmid DNA is cleaved first to open circular DNA, which in turn is converted to linear DNA and then to fragments. It was found that both aromatic ring and carboxyl group in the arene structure are needed for effective cleavage reaction. Orthobrombenzoic acid, parabrombenzoic acid, orthochlorbenzoic acid, parachlorbenzoic acid and salicylic acid were found to efficiently cleave DNA in the presence of Cu(II) at concentration as low as 0,01mM. The cleavage was dependent on the presence of O_2. Catalase and sodium azide, scavengers of oxygen-derived species, inhibited DNA cleavage, suggesting that both superoxide radical and singlet oxygen may be involved in the cleavage mechanism. The DNA cleavage efficiency of the designed system arene-Cu is comparable to that of the well-known DNA cleaving reagents such as phenanthroline-Cu and ascorbic acid-Cu. However in contrast to the mentioned reagents, the system arene-Cu does not require external reducing agents or H_2O_2.

Introduction

A number of metal complexes capable of inducing DNA damage have been developed. Bleomycin-iron, phenanthroline-Cu, ascorbic acid - Cu and few related copper complexes act like catalysts causing oxidative damage to nucleic acids [Sigman, 1993; Stubbe, 1987]. Some of metal complexes were used as reactive groups in oligonucleotide conjugates designed for cleaving nucleic acids. These conjugates were shown to cleave complementary DNA, although with low efficacy and in conditions far from physiological conditions, usually in the presence of a reducing agent or H_2O_2. Sequence-directed cleavage of DNA has been achieved by means of complementary oligonucleotides bearing terminal reactive group. Some metal complex systems, for instance, ascorbic acid-Cu are inexperienced to conjugate with oligonucleotides. The goal of our study is the development of chemical cleavage systems that are easily amenable to chemical synthetic manipulation and have suitable biocompatability. We propose to use as DNA-cleaving groups complexes of arenes with Cu. These complexes are unique in that arenes are small simple compounds and can be easily modified to change their reactivity. A variety of simple and rather inert arenes that react with nucleophiles in the presence of specific metal ions can be easily conjugated to oligonucleotides and peptides.

Methods

All experiments were carried out with plasmid mdr1cat5 [Uchiumi, 1993]. Plasmid DNA (400 ng) was incubated at 37^0C with arenes plus $CuSO_4$ in 10 mM-sodium phosphate buffer, pH 7,5. Specific details of incubation period and arene and Cu(II) ion concentrations are given in legends to Figures. Quenching of reactions was achieved by adding Na_2EDTA to 2mM concentration and 1/10 volumeof loading buffer (0,25% bromphenol blue, 0,25% xylene cyanol, 80% glycerol in 10 mM Tris-HCl, pH 7,5) was added to each tube. Samples were loaded onto a 1% agarose gel containing ethidium bromide (1μg/mL). Electrophoresis was carried out for 2 h at 120 V in TBE buffer (50 mM Tris-borate, pH 8,4, 1 mM EDTA) and visualized by UV illumination.

Results and discussion

We have investigated cleavage of plasmid DNA by a simple arene-Cu system under physiological conditions. In order to assess the competence of orthobrombenzoic acid (**1**), orthochlorbenzoic acid (**2**), parabrombenzoic

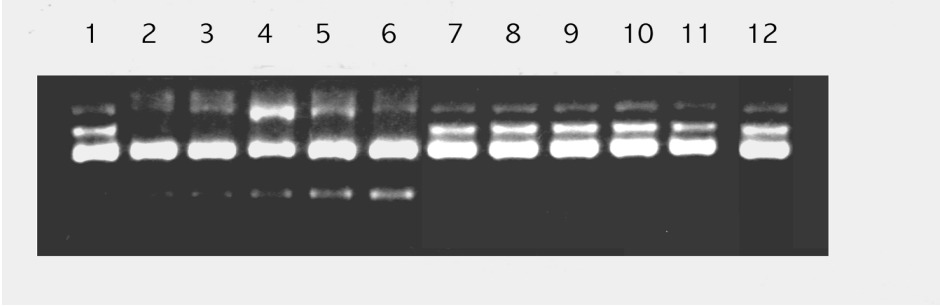


Figure 1. Agarose gel analysis of plasmid DNA cleaved by compounds 1 – 5.

Plasmid DNA was incubated at 370C for 1h with or without 10 μM compounds 1 –5 and 10 μM CuSO4. Lane 1- DNA without reagents; lane 2-6 DNA incubated with compounds 1-5 respectively; lane 7-11 DNA incubated with compounds 1-5 respectively without CuSO4; lane 12 – plasmid DNA incubated with CuSO4.

acid (**3**), parachlorbenzoic acid (**4**) and salicylic acid (**5**) for DNA strand scission, each of the compounds was incubated with DNA under identical reaction conditions. We have found that the compounds cleave DNA in the presence of Cu(II) at concentration as low as 10 μM (lane 2, fig1). In the absence of Cu(II) ions , arenes are inert to the DNA (lane 7-11, fig 1) up to 10 mM concentration. Figure 1 show that all the compounds display similar DNA cleaving reactivity.

To identify the groups in the structure of the tested arenes that are essential for the cleavage reaction, we tested the effect on DNA of compounds 1-5 and phenol, benzoic acid, brombenzol and sodium acetate. Only benzoic acid cleaved DNA as well as 1-5 (date not shown). These results suggest that aromatic ring and carboxyl group are both necessary for effective cleavage reaction, when π and σ coordination system is present. We have found that all the investigated compounds can cleave plasmid DNA to some extent without addition of Cu(II) if their concentrations are higher then 0, 01M. An explanation could be that in this case the natural plasmid bound copper ions participated in the reactions, because addition of Na_2EDTA strongly inhibited cleavage of DNA. This fact indicates that Cu(II) and arenes both are essential for DNA damaging reaction. No cleavage occurred when the plasmid DNA was incubated with Cu(II) at mM in the absence of the compounds under investigation.

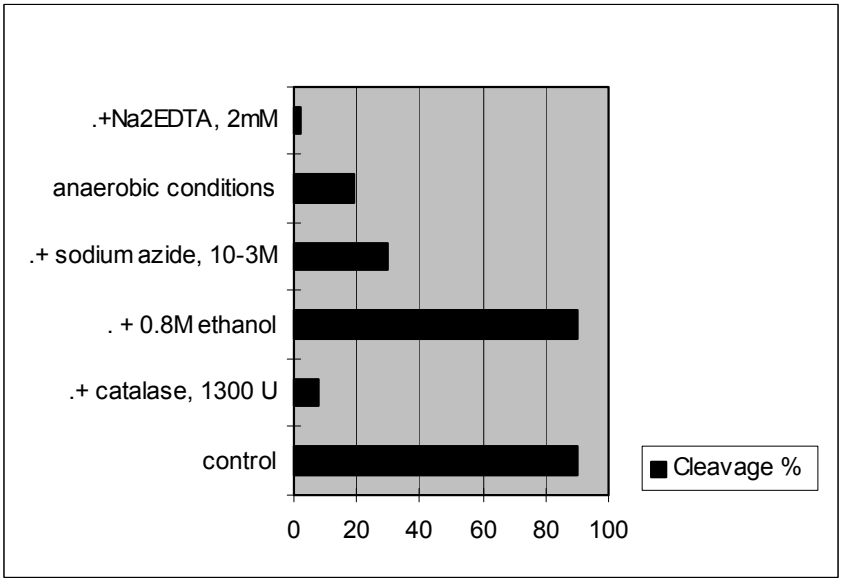


Figure 2. Plasmid DNA cleavage by orthobrombenzoic acid under anaerobic conditions and with oxygen species scavangers.

Cleavage reactions were carried out for 3h at 370C. Reaction mixtures contained 10 μM orthobrombenzoic acid, 10 μM CuSO4 and other components as indicated in the figure. To provide anaerobic condition all mixture components were degassed for 20 min by argon before the reaction.

To elucidate the mechanism of the reaction we have investigated potential role of oxygen in the process of DNA cleavage. The reaction was performed under normal atmospheric conditions, under anaerobic conditions and in the presence of various reactive oxygen species scavengers. The data presented in figure 2 demonstrate that cleavage by Cu(II)-arylhologenid does not occur under anaerobic condition. The obvious role of O_2 might be in the generation of oxygen radicals such as $^{\cdot}OH$, which could be the actual mediators of DNA destruction [Cohen, 1983; Pogozelsski, 1998]. As shown in figure 2, singlet oxygen scavenger, sodium azide is also much more

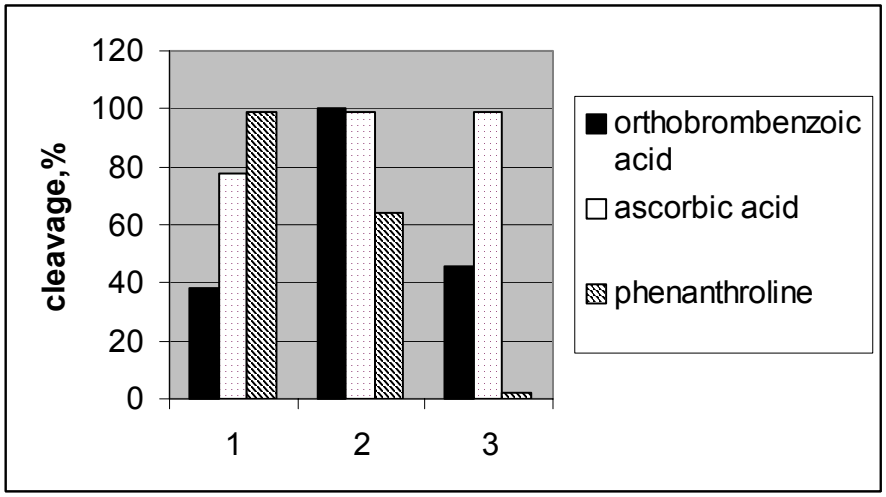


Figure 3. Cleavage of plasmid DNA by metal complexes. Effect of □-mercaptoethanol and H2O2. Cleavage reaction were carried out for 2h at 370C. Reaction mixtures contained 10 μM CuSO4, 10 μM reagents, and 1mM 2-mercaptoetanol or 0,01mM H2O2. 1- cleavage with 2-mercaptoetanol, 2 – cleavage with H2O2, 3 – cleavage without additional factor.

effective as an inhibitor than the ·OH radical's scavenger, alcohol. Additional evidence for oxygen dependence of the cleavage in the presence of arenes and Cu(II) ions was obtained in the experiment with catalase, an enzyme that disproportionate H_2O_2 to yield H_2O+O_2. The enzyme completely inhibits the cleavage (Fig.2). Therefore it can be concluded, that both H_2O_2 and singlet oxygen participate in the reaction.

Generation of active oxygen species can occur as a result of changes of oxidizing state Cu^{II}/Cu^{I} or Cu^{II}/Cu^{III} [Yamomoto, 1989; Thederahn, 1989]. To find out, what active species participate in the DNA damage, the effect of reducing or oxidizing agents H_2O_2 and β-merkaptoethanol on the DNA cleavage induced by Cu(II)-arenes was investigated. As shown in figure 3, addition of H_2O_2 stimulates the cleavage, while addition of β-mercaptoethanol does not effect the process.

This result indicates that the DNA cleavage reaction by arenes with Cu^{II} is more likely accompanied by transition of the cooper oxidizing effect Cu^{II}/Cu^{III} then Cu^{II}/Cu^{II}. We have compared the cleavage efficacy of the system orthobrombenzoic acid-Cu with those of the well-known DNA cleaving reagents: phenanthroline-Cu [Chen, 1986] and ascorbic acid-Cu [Drouin, 1996; Wang,1989]. Figure 3 demonstrates that the efficiency of DNA cleavage without additional agent in the system orthobrombenzoic acid-Cu is comparable with that for ascorbic acid-Cu. Under identical condition cleavage does not occure in the system phenanthroline-Cu without addition of β–mercaptoethanol.

We have demonstrated that arenes of different structure containing carboxyl groups can effectively cause cleavage of DNA in the presence of Cu under physiological condition. These reagents can be used for the design of a new class of oligonucleotide conjugates, affinity reagents for inactivating viral nucleic acids and specific oncogenes.

The study is supported by the grant INTAS No. 99-630.

References

1. Chen C.-H. B., Sigman D.S. (1986) Nuclease activity of 1-10-phenantroline-copper. Sequence-specific targeting. *Proc. Natl. Acad. Sci. USA.* **83**, 7147-7151.
2. Cohen G., Greenwald R.A.(1983) Oxy radicals and their scavenger systems. New York. (Eds.).
3. Drouin R., Rodriguez H., Gao S. W., Gebreyes Z., O'Connor T. R., Holmquist G. P. & Akman S. A. (1996) Cupric ion/ascorbate/hydrogen peroxide-induced base modifications. *Free Radic. Biol. Med.* **21**, 261-263.
4. Knorre D.G., Vlassov V.V., Zarytova V.F., Lebedev A.V., Fedorova O.S. (1994) Design and targeted reactions of oligonucleotide derivatives. Boca Raton: CRC Press.
5. Pogozelski W. K. & Tullius T. D.(1998) Oxidative strand Scission of nucleic acids: routes initiated by hydrogen abstraction from the sugar moiety. *Chemical reviews* **98**, 1089-1108.
6. Sigman D. S., Mazumder A. & Perrin D. M. (1993) Chemical nucleases. *Chemical reviews.* **93**, 2295-2316.
7. Stubbe J., Kozarich J. W. (1987) Mecanizms of bleomycin-induced DNA degradation. *Chemical reviews.* **87**, 1107-1136.
8. Thederahn T. B., Kuwabara M. D., Larsen T. A. & Sigman D. S. (1989) Nuclease activity of 1,10-phenantroline-copper: kinetic mechanism. *J. Am. Chem. Soc.* **111,** 4941-4946.
9. Uchiumi T., Kohno K., Tanimura H., Hidaka K., Asakuno K., Abe H., Uchida Y. and Kuwano M. (1993) Involvment of protein kinase in environmental stress-indused activation of human multidrug resistance 1 (MDR1) gene promoter. FEBS, **326**, 11-16.
10. Wang Y. & Ness V. (1989) Site-specific cleavage of supercoiled DNA by ascorbate/Cu(II). *Nucl. Acids Res.* **17**, 6915-6926.
11. Yamomoto K., Kawanishi S.(1989) Hydroxyl radical is not the main active species in the site-specific DNA damage induced by Cu(II) ion and hydrogen peroxide. *J. Biol. Chem.* **264**, 15435-15440.

BEHAVIORAL ANALYSIS OF HOMEOTIC LAWC-MUTANTS IN *D. MELANOGASTER*

Simonova O., Sukhoverkhova T., Modestova E., *Korochkin L.

Lab. neurogenetics and developmental genetics, Institute of Gene Biology, Moscow, Russia
e-mail: likor@mx.ibg.relarn.ru
*Corresponding author

Keywords: mutation, leg-arista-wing-complex

Resume

An earlier obtained regulatory mutation of D. melanogaster named as leg-arista-wing-complex (lawc) causes a weak homeotic arista transformation into tarsus. In order to check how homeotic transformation influences fly behavior, we studied the olfactory and proboscis-extension responses, since taste and odor chemoreceptors are located on foreleg tarsi and on antennae respectfully. Despite low penetrance of transformed aristae, the attraction index of lawc-mutants appears to be less than that of control wild type flies. During taste experiments the conclusion was made, that homeotic legs have a positive response on sugar stimulation. Afferent projections of sensory neurons were examined after cobalt chloride impregnation. We found that these

Neural routes starting from homeotic legs reach thoracic ganglia. The date of genetic analysis demonstrates the interaction between lawc and other proneural loci: achaete, scute, cut. Northern blot analysis revealed 11 poly A+ mRNAs in all stages of development. The largest transcript was about 12 kb in length.

The study is supported by the grant INTAS No. 96-1652.

COMPLEX FORMATION OF Fpg PROTEIN WITH OLIGONUCLEOTIDES AND THEIR DERIVATIVES

Ishchenko A.A., *Nevinsky G.A.

Institute of Bioorganic Chemistry SB RAS, Novosibirsk, Russia
e-mail: nevinsky@niboch.nsc.ru
*Corresponding author

Keywords: Fpg protein, 8-oxoguanine-DNA-glycosylase, substrate specificity, inhibition, footprinting

Resume

Motivation & Results:

A variety of mutagenic and cytotoxic DNA lesions are formed by oxidative DNA damage, one of the most abundant beings 8-oxoguanine (oxoG). The affinity to Fpg protein for oligonucleotides (ON) $d(pC)_n$, $d(pT)_n$ and $d(pA)_n$ and their duplexes was studied. All single- (ss) and double stranded (ds) ONs were competitive inhibitors of the reaction catalyzed by this glycosylase. The K_i values for different ON and their complexes were determined. Lengthening of $d(pN)_n$ and their duplexes by one unit increases the affinity to glycosylase 1.4-1.6 and 2.2 times, respectively. The data obtained indicate that high affinity of Fpg for DNA is provided mainly due to additivity of many weak interactions of the enzyme with 10 and 14 units of sugar-phosphate backbone of ss and ds ONs, respectively. Transition from nonspecific to specific ONs containing oxoG leads to enhancement of their affinity to the enzyme by a factor of ~10, while the rate of the reaction increases by 6-8 orders of magnitude. The DNA-Fpg complex formation was shown cannot provide for enzyme specificity. Specificity of enzyme action is provided by the stages of the enzyme-dependent DNA adaptation to the optimal conformation and directly of catalysis.

Introduction

Genomic DNA is susceptible to attack by reactive oxygen species produced by cellular aerobic metabolism and by exogenous agents, including certain chemical carcinogens and ionizing radiation. Fpg removes oxoG and other purine lesions from DNA, including imidazole ring-opened guanine and imidazole ring-opened adenine [1]. OxoG is both mis-coding and mutagenic, and is believed to be the major physiological substrate for activity, which cleaves both the 3'- and 5'-phosphodiester bonds at an AP site by successive β- and δ-eliminations, leaving both the 3'- and 5'-ends of the gap phosphorylated [2]. Several nuclear and itochondrial activities recognizing oxoG have been identified in mammalian cells. The chemical nature of the modified base is not the only parameter for the recognition and/or excision by Fpg. Although a *fapy* residue is a substrate for Fpg in B-DNA, it is not repaired when encountered in Z-DNA [3].

Methods

The sequences of 23-mer ONs used in this work are listed below.

OG1, oxoGCTCTCCCTTCCTCCTTTCCTCT;

OG2, C-oxoGTCTCCCTTCCTCCTTTCCTCT;

OG11, CTCTCCCTTC-oxoGCTCCTTTCCTCT;

ONs containing the oxoG base in various positions are marked as OG**N**, where *n* is the distance from the ON 5'-end to the oxoG residue. Oligonucleotides containing unmodified G, instead of oxoG, are marked in the text as G1-G11. The 23-mers containing oxoG were [5'-^{32}P]-labeled and annealed to the complementary sequences possessing dC positioned opposite the oxoG residue. These complementary ONs are marked as C**N** (C1-C11), where *n* is the number of oxoG (or non-modified dG) base in the complementary strand.

Results and discussion

As shown earlier for number of DNA-dependent enzymes (for review see [4]) interact effectively not only with specific ss and ds ONs, but also with non-specific ones. Therefore in this article we examined whether Fpg can also bind to short nonspecific ONs. All specific and non-specific ss and ds ONs were found to be competitive inhibitors of the enzyme. Under the conditions used, when the concentration of a ds OG3 substrate is $2K_M$, I_{50} should be equal to $3K_i$ [5], and just this was observed in experiment. The I_{50} values with the account of $I_{50} = 3K_i$ ratio were used as quantitative characteristics of the ON affinity to the enzyme. The I_{50} values of model ss- and ds $d(pN)_n$ are summarized and shown in Fig. 1 as lg I_{50} dependencies on the ligand length (*n*).

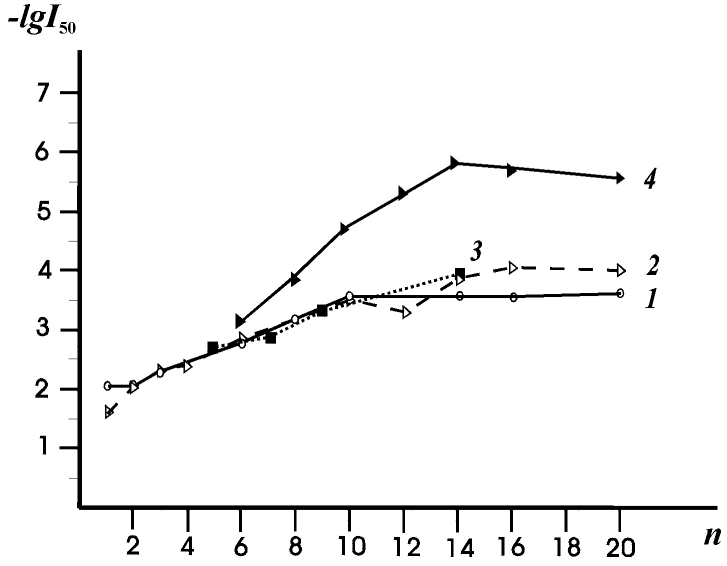


Figure 1. Dependence of –lg values of I_{50} on the $d(pN)_n$ length (n). Curves 1-4 correspond to $d(pA)_n$, $d(pT)_n$, $d(pC)_n$, and $d(pA)_n \bullet d(pT)_n$ duplex, respectively

It should be mentioned that all dependencies are linear up to n = 9-10 and n = 13-14 for ss- and ds ONs, respectively. The linearity is indicative of an additive interaction of individual nucleotide units of ss ONs and nucleotide pairs of ds ONs with Fpg.

Fig. 1 demonstrates that the affinity of Fpg for ss ONs does not practically change at $n > 10$, whereas the affinity of duplexes changes at $n \leq 14$. This may point to a change in the number of nucleotide units within the protein globule upon transition from ss- to ds ONs. To check this assumption, we determined the number of nucleotide units protected by the enzyme against nuclease hydrolysis within Fpg complexes with ss- and ds ONs.

DNA protection from nuclease hydrolysis. The enzyme complexes with ss and ds ONs were cross-linked to one ON strand using *cis*-aquahydroxydiamminoplatinum [6]. The results of electrophoretic analysis of enzyme-protected ON are showed in Fig. 2. It is seen that the enzyme protected from DNase hydrolysis of 9-10 and 13-14 links of ss- and ds ON, respectively, that completely correlates with those deduced from logarithmic dependencies (Fig. 1).

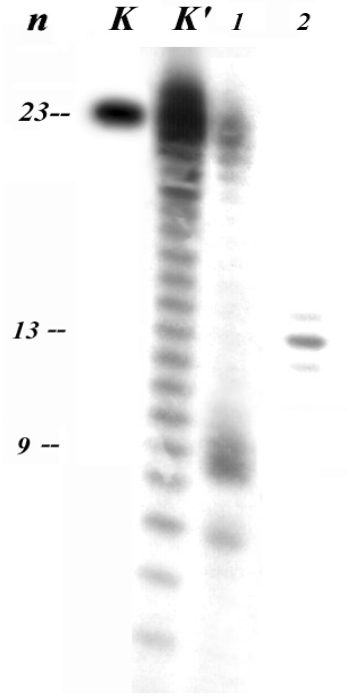


Figure 2. Electrophoretic analysis of lengths of nonspecific ss and ds ONs protected by Fpg against hydrolysis by nucleases. K – initial G11 oligonucleotide modified by Pt^{2+} reagent, K'- as K but after cleavage of the modified G11 by the method of Maxam and Gilbert. 1 and 2, the adducts of modified ss (1) and dsG11 (2) with Fpg after their treatment by DNase and NaCN consequently.

The results of the analysis of the intermediate (Schiff base) formation of the enzyme during its AP-lyase activity were used to determine the localization of the AP-endonuclease center. It was found that the 5'-terminal [^{32}P]-label of OG1-OG4 oligonucleotides after reduction of their intermediates with the enzyme using $NaBH_4$ is protected by the enzyme against nuclease and phosphatase cleavage (data not shown). A noticeable decrease of the label in the protein adducts with ONs is observed only with OG5 - OG11, regardless of their form.

Thus, it can be concluded that the AP-lyase center of the enzyme is localized in the 5'-terminal region of the DNA-binding site at a distance not exceeding 4 nucleotides from the 5' end of 9-13-mer ONs bound by the enzyme.

Enzyme interaction with nonspecific ONs and mononucleotides. Based on the slopes in Fig. 1, the changes in K_i (K_d) for all ss ONs studied can be described by a descending geometric progression:

$K_d[(dpN)_n] = K_d[(dNMP)] \bullet [K_d(n)]^{n-1} = K_d[(dNMP)_n][1/f]^{n-1}$

where $K_d[(dNMP)]$ is the K_d value characterizing enzyme affinity to its minimal ligand (dNMP), $K_d(n) = 1/f - K_i$ for any nucleotide unit of extended ligand (at $1 < n < 10$ these values are the same); n is the number of nucleotide units in the $d(pN)_n$, f is the factor of affinity increase at ligand lengthening by one nucleotide unit.

Analysis of the relative contribution of dNMP structural elements (minimal ligand) to its affinity to glycosylase active center is of special interest. As follows from the table, the I_{50} values (25-30 mM) for orthophosphate, 5'-deoxyribosephosphate (pdR), and dNMP are practically the same, while I_{50} for deoxyribose (dR) is higher than 0.4 M. Thus, the phosphate group of these ligands makes the main contribution to recognition by the enzyme of dNMP or one nucleotide unit of lengthy ONs.

Table 1.

Nucleotides and their components	I_{50}, mM	Non-specific ONs	I_{50}, μM	Specific Ons	I_{50}, μM
KH_2PO_4	**30**	G11	**30**	OG11	**7**
dR	**470**	G6	**30**	OG6	**6**
pdR	**25**	C11	**18**	OG1	**6**
dAMP	**10**	C3	**30**	OG2	**1**
dTMP	**25**	$d(pF)_9pT$	**250**		
dCMP	**25**	$d(pT)_{10}$	**350**	OG1•C1	**2**
dGMP	**10**	G11•C11	**200**	OG2•C2	**3**

Error in I_{50} determination was 20-40%. F - tetrahydrofuran derivative.

The 1.4-1.6-fold increase in affinity upon ss ON elongation by one nucleotide unit corresponds to a change in the Gibbs free energy by -0.2 to -0.28 kcal/mole. This is comparable to the corresponding values for weak hydrophobic, ion-dipole, and dipole-dipole interactions [4].

As shown by us earlier [4], weak nonspecific interactions of individual ON's links with enzymes may be uniform and consist in superposition of several types of weak contacts of different structural element of DNA monomers with the enzyme.

The f values (1.40-1.60) and K_d (0.62-0.71 M) characterizing the Fpg weak interactions with each nucleotide unit of ss ON are practically independent of the link nature. This is indicative of the decisive contribution of the ss ON sugar-phosphate backbone into their recognition by Fpg. The affinity of all $d(pN)_{10}$ to the enzyme practically does not differ from that of abasic decamer $d(pF)_9pT$ (Table) containing hydrogen atom instead bases, which is also confirm a very important role of the sugar-phosphate backbone in providing of high affinity of Fpg for ss and ds DNA.

As seen from Fig. 1, there is no essential stabilization of short duplexes is achieved by their interaction with Fpg (as is the case of some other enzymes [4]). The change in duplex affinity upon their elongation, like $(dpT)_n$ - $(dpA)_n$, $n \geq 8$ is described by a descending geometric progression as for ss ON: the factor f is about 2.2. This speaks in favor of the enzyme interaction with both strands of ds DNA, but the contribution of the second strand is essentially lower than that of the first.

As can be seen in the Table, substitution of dG by 8-oxodG in ss ON increases its affinity to the enzyme only by a factor of 4-5. As for duplexes corresponding to OG1 and OG2 (bad substrates [6]), and duplexes containing dG instead of 8-oxodG, the difference in the affinity is 7- to 10-fold. Thus, the contribution of the specific 8-oxodG nucleotide to the affinity of modified DNA to Fpg does not exceed one-two orders of magnitude. At the same time, the sum of many weak additive interactions of DNA with the glycosylase provides about 7 orders of the affinity. As follows from the literature [1] and our data, Fpg is not able to excise bases from non-modified DNA with noticeable efficiency: the rate of nonspecific enzyme action decreases by 6-8 orders. Thus, contrary to the common notion, the stage of complex formation cannot provide for Fpg specificity. Taken together, it is obviously that like in the case a number of sequence-dependent repair, topoisomerization, recombination, integration, and restriction enzymes [4] specificity of Fpg action is provided by the stages of the enzyme-dependent DNA adaptation to the optimal conformation and directly of catalysis: k_{cat} increases by 4-8 orders of magnitude when changing from the nonspecific to specific DNA.

Acknowledgements

We are deeply grateful to Prof. Jacques Laval (the Gustave Roussy Institute, France) for the gift of the *E.coli* strain used for the preparation of Fpg. The research was made possible in part by a grant from INTAS No. 99-

1645, from The Welcome Trust (England), from the Russian Foundation of Basic Research and a grant from the Siberian Division of Russian Academy of Sciences.

References

1. Tchou,J., Bodepudi,V., Shibutani,S., Antoshechkin,I., Miller,J., Grollman,A.P. and Johnson,F.J. (1994) Substrate specificity of Fpg protein. Recognition and cleavage of oxidatively damaged DNA. *Biol. Chem.,* **269**, 15318-15324.
2. Bailly,V., Verly,W.G., O'Connor,T., Laval,J. (1989) Mechanism of DNA strand nicking at apurinic/apyrimidinic sites by *Escherichia coli* [formamidopyrimidine]DNA glycosylase. *Biochem. J.* **262**, 581-589.
3. Kouchakdjian,M., Bodepudi,V., Shibutani,S., Eisenberg,M., Johnson,P., Grollman,A.P. and Patel,D.J. (1991) NMR structural studies of the ionizing radiation adduct 7-hydro-8-oxodeoxyguanosine (8-oxo-7H-dG) opposite deoxyadenosine in a DNA duplex. 8-Oxo-7H-dG(syn).dA(anti) alignment at lesion site. *Biochemistry* **30**, 1403-1412.
4. Nevinskii, G.A. (1995) An important role for weak interactions during the recognition of long DNA and RNA molecules by enzymes. *Mol. Biol. (Moscow)* **29**, 16-37.
5. Kelleti,T., Bases of Enzyme Kinetics.Translated under the title Osnovy fermentativnoi kinetiki., Moscow: Mir, 1990, pp. 133-138.
6. Ishchenko,A.A., Bulychev,N.V., Maksakova,G.A., Johnson,F., Nevinsky,G.A. (1998) Interaction of *Escherichia coli* 8-oxoguanine DNA glycosylase with single-stranded oligodeoxyribonucleotides and their complexes, *Mol. Biol.* (Moscow), **32**, 454-462.

EFFICIENT SITE-DIRECTED CLEAVAGE OF YAST tRNA^{PHE} WITH OLIGONUCLEOTIDES CONJUGATED TO IMIDAZOLE CONSTRUCTS

Beloglazova N.G., [1]Polushin N.N., Sil'nikov V.N., Zenkova M.A., Vlassov V.V.

Institute of Bioorganic Chemistry SB RAS, Novosibirsk, Russia
e-mail: nnn@niboch.nsc.ru
[1]Fidelity Systems, Inc., 7961 Cessna Ave., Gaithersburg, MD 20879, USA
*Corresponding author

Keywords: sequence-specific RNA cleavage, antisense oligonucleotides, tRNA

Resume

New RNA cleaving dendrimer constructs, containing two or four imidazole groups have been designed. The groups were conjugated to 5′ -end base of antisence oligonucleotides complementary to the 3′ end of tRNA^{Phe}. The synthesized conjugates efficiently cleaved the tRNA within the target sequence adjacent to the oligonucleotide-binding site.

Motivation:

Development of oligonucleotide derivatives capable of cleaving RNA at complementary sequences can provide new tools for investigation of RNA structure and functions and can open new possibilities for design of antisense therapeutics. A straightforward approach to design of RNA cleaving compounds consists in mimicking of active centers of natural ribonucleases by conjugation of organic constructs containing reactive groups that usually are constituents of the active centers of nucleases to molecules possessing affinity to RNA. Covalent attachment of RNA cleaving constructs to oligonucleotides can provide conjugates capable of cleaving RNA targets in a sequence-specific manner [Trawick et al., 1998; Vlassov et al., 1997; Haner et al., 1997].

Here we describe novel oligonucleotide conjugates, bearing catalytic RNA cleaving groups with two or four imidazole residues. The conjugates were obtained by functionalization of oligonucleotide using the precursor approach, in the course of standard solid-phase synthesis. Ribonuclease activity of the conjugates was tested using yeast tRNA^{Phe} as a model substrate. It was shown that the designed conjugates hydrolysis tRNA^{Phe} with high efficiency at the target site.

Methods

Yeast tRNA^{Phe} was a generous gift of Dr. G. Keith. Oligonucleotide pATCGAACACAGGACCT (B) was synthesized on ASM-700 DNA Synthesizer (Biosset, Russia) by standard solid-phase phosphoramidite procedure. Functionalized oligonucleotides were synthesized using strategy of precursor chemistry. RNA cleaving constructs 1 and 2 were prepared as described in [Polushin] 3′-[^{32}P]-tRNA^{Phe} was obtained according to published protocol [Ehresmann et al., 1987]. Cleavage of tRNA^{Phe} with oligonucleotide conjugates B1 and B2 was carried out at 37°C. Reaction mixture (10 μl) contained 50 imidazole buffer, pH 7.0, 200 mM KCl, 1 mM EDTA, 100 μg/ml *E. coli* total tRNA as carrier, 5x10^{-7} M 3′-[^{32}P]-tRNA^{Phe} and oligonucleotide conjugates B1, B2 at concentration of 1x10^{-5} M. Reactions were quenched by precipitation of tRNA and tRNA fragments with 150 μl of 2% lithium perchlorate solution in acetone. tRNA cleavage products were analyzed by electrophoresis in 12% denaturing PAAG. To identify cleavage sites, imidazole ladder and partial RNase T1 digest of tRNA were run in parallel. To obtain quantitative data, Cherenkov's counting dried gels. Total extent of tRNA cleavage was determined as ratio of radioactivity measured in each tRNA fragment to total radioactivity applied the gel.

Results and discussion

Bis- and tetraimidazole containing constructs (Fig.1) were attached to the 5'-end nucleobase of oligonucleotide pATCGAACACAGGACCT (B) in the course of synthesis. The oligonucleotide is complementary to nucleotides 44-60 within the 3′-part of yeast tRNA^{Phe} (Fig. 2). Earlier it was shown that this oligonucleotide could efficiently bind to the target sequence in the tRNA^{Phe} [Petyuk et al., 1999]. 75-85 % complex formation is achieved at 5x10^{-5} M concentration of oligonucleotide B. The conjugates were constructed so as to deliver the catalytic groups to the two adjacent CA sequences in the T-arm of the tRNA (positions 61-62 and 63-64). Enhanced susceptibility of the 5′-Py-A-3′ sequences in RNA to RNase A and ribonuclease mimics is well known [Komiyama et al., 1994]. Fig. 3 displays a typical autoradiograph of the tRNA^{Phe} cleavage pattern by the conjugates B1 and B2. Both conjugates cleave the tRNA with similar specificity: at C63-A64 and A-64-G65. The cleavages occur 3

bases apart from the 5'-terminal nucleotide C60 complexes with oligonucleotide. Similar specificity was observed in experiments with conjugates with different cleaving groups and linkers. Positions of the cleavage sites are consistent with electrophoretic mobilities of the degradation products, which migrate like oligonucleotides generated by RNase T1 or by imidazole. Rates of RNA cleavage by reagents B1 and B2 are different (Fig.3 and Fig. 4). Conjugate B2 with four imidazoles cleaves the tRNA approximately three times faster as compared to B1. Complete cleavage of the tRNA is achieved by 1.5 and 5 hours of incubation with reagents B2 and B1 (10 μM), respectively.

The oligonucleotide conjugate described in the present paper demonstrates high cleavage efficiency. This efficiency is achieved by design of the structure of the RNA cleaving group and by choice of the target RNA sequence that is particularly sensitive to RNase mimics. Further investigation of factors affecting susceptibility of phosphodiester bonds to hydrolyzing agents and optimization of structure of the catalytic group will lead to design of efficient RNA cleaving oligonucleotide conjugates that will find applications in development of antisense therapeutics.

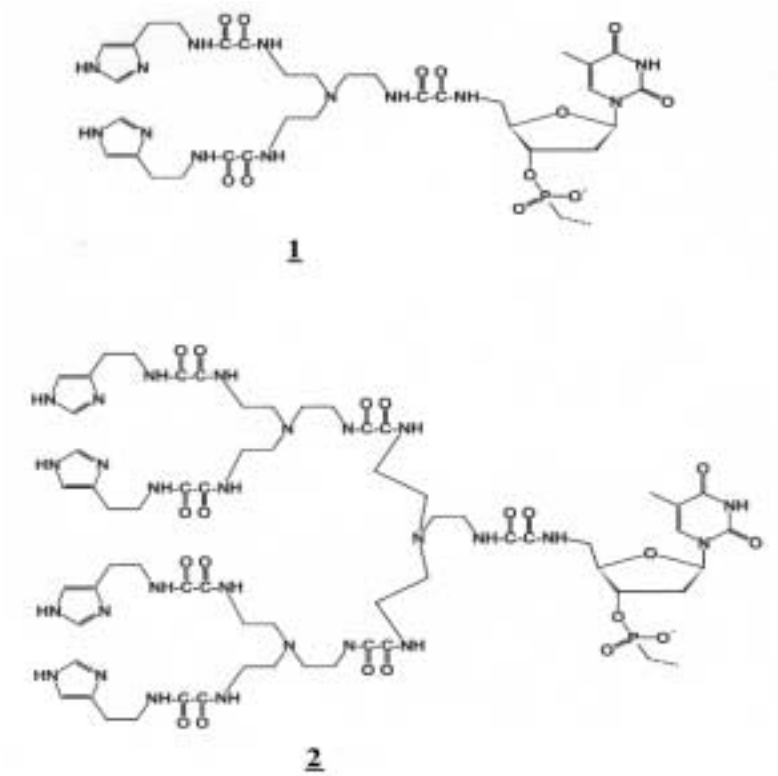


Figure 1. The catalytic constructs attached to 5′-end nucleobase of oligonucleotide B.

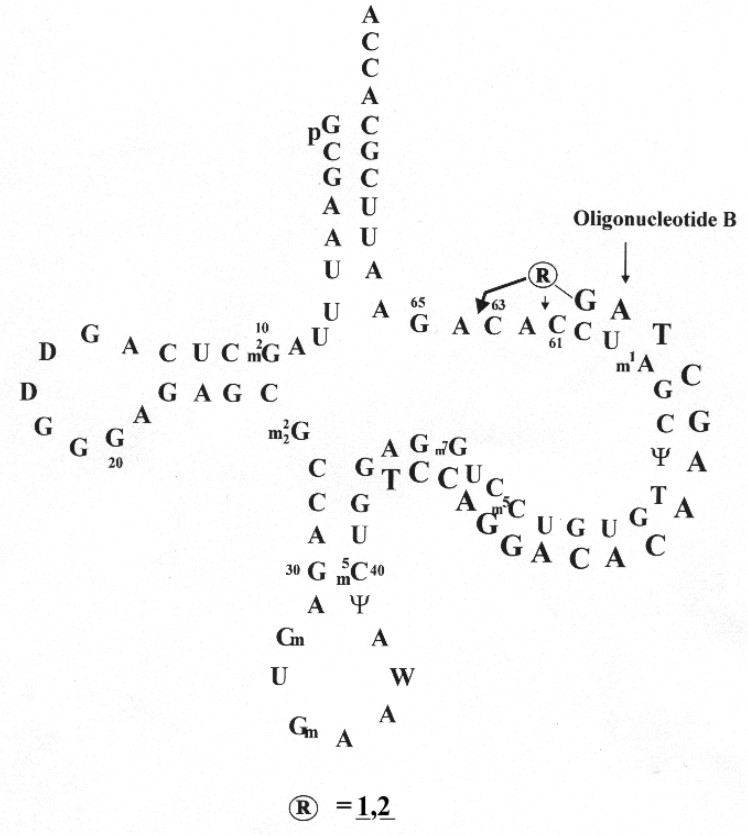


Figure 2. tRNA^{Phe} with complementary oligonucleotide B. R-catalytic group attached to 5$^{/}$ -end base of the oligonucleotide. Arrows indicate phosphodiester bonds attacked by the catalytic group.

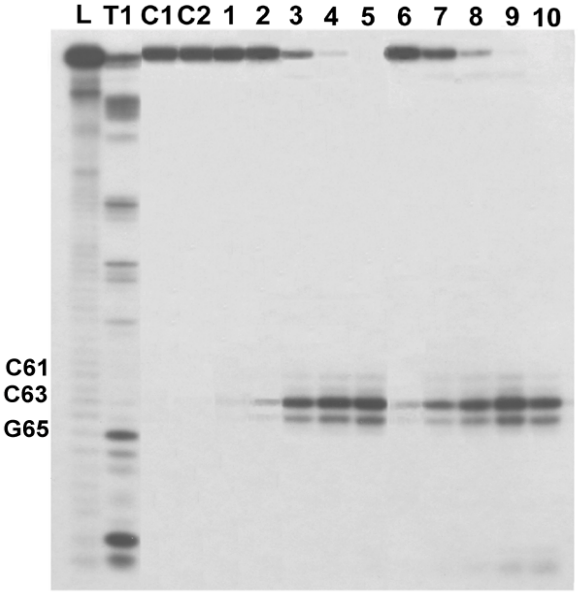


Figure 3. Cleavage of tRNA ^{Phe} with oligonucleotide conjugates B1 and B2. The reaction was performed at 37°C in 50 mM imidazole buffer, pH 7.0, containing 200 mM KCl, 1 mM EDTA, 100 μg/ml total tRNA as a carrier, 5x10^{-7} M 3$^{/}$- [^{32}P] tRNA ^{Phe} and 1x10^{-5} M B1 and B2.

L - partial imidazole ladder;

T1 - partial digest with ribonuclease T1;

C1 - tRNA ^{Phe} incubated in the reaction buffer at 37°C for 5h;

C2 - tRNA ^{Phe} incubated at 37°C for 5h in the reaction buffer in the presence of 5x10^{-5} M oligonucleotide B;

1-5 - incubation of tRNA ^{Phe} with conjugate B1 for 0.5h, 1h, 3h, 5h, 8h, respectively;

6-10 - incubation of tRNA ^{Phe} with compound B2 for 0.5h, 1h, 3h, 5h, 8h, respectively.

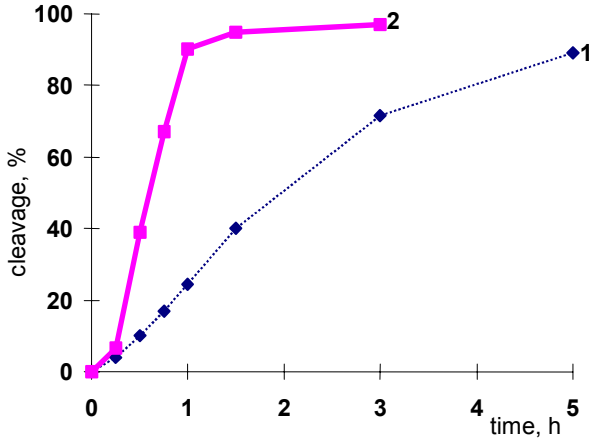


Fig. 4. Time course of $tRNA^{Phe}$ cleavage by oligonucleotide conjugates B1 (curve 1) and B2 (curve 2). The reaction was performed at 37°C in 50 mM imidazole buffer, pH 7.0, containing 200 mM KCl, 1 mM EDTA, 100 μg/ml total yeast tRNA as carrier, $5x10^{-7}$ M 3′-[^{32}P]-$tRNA^{Phe}$. Concentration of the conjugates was $1x10^{-5}$ M.

The study is supported by the grant INTAS.

References

1. B.N. Trawick, A.T. Daniher, J.K. Bashkin, Chem. Rev. **V**, 98. P. 939-960(1998).
2. V. Vlassov, T. Abramova, T. Godovikova, R. Gige, V. Silnikov, Antisense and Nucl Acid Drug Dev. 7. P. 39-42(1997)
3. R. Haner,. and J. Hall, Antisense and Nucl. Acid Drug Dev. *7*, 423-430(1997)
4. N.N. Polushin, "The Precursor Strategy: Terminous Methoxyoxalamido Modifiers for Single and Multiple Oligonucleotide Tethering". Manuscript submitted
5. C. Ehresmann, F. Baudin, M. Mougel, P. Romby, B. Ehresmann, J.-P. Ebel, Nucleic Acids Res. 15. P. 9109-9128(1987)
6. V. Petyuk, M. Zenkova, R. Giege, V. Vlassov, FEBS Letters 444, 217-221(1999)
7. M. Komiyama, T. Inokava, J. Biochem. 116, 719-720(1994).

COMPARATIVE CHARACTERIZATION OF THE 5'-REGIONS OF THE VOLE, HUMAN AND MOUSE XIST GENE: CONSERVED AND REPETITIVE SEQUENCES

Kolesnikov N.N., Nesterova T.B., Slobodyanyuk S.Y., Shevchenko A.V., Pavlova M.V., Rogozin I.B., Elisaphenko E.A., Kel A.E., Levitsky V.G., Brockdorff N., Zakian S.M.

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: kolesnik@bionet.nsc.ru
*Corresponding author

Keywords: Xist, X-inactivation, repeats, SINE

Resume

To identify the conserved homologous sequences, which may be important in the Xist regulation, 5' sequences upstream of the gene in four vole species were cloned, sequenced and compared with the corresponding mouse sequence. Comparative analysis demonstrates the high level of homology (more than 90%) between four vole species. The differences between vole species are accounted mainly for presence or absence some repetitive sequences, small insertions and deletions, transitions and transversions.

Comparison between vole and mouse Xist upstream regions has not revealed any significant homology; overall identity has not exceeded 50%. However, three regions with much higher level of homology were detected (79%, 66% and 65% of identity). First region corresponds to minimal promoter sequence, from –1 to –101 [Hendrich et al., 1997], second sequence is located upstream of P1 promoter and about 400 bp in size. It has homology 61% with appropriate sequence of horse and 60%with human. Inside of this sequence we found three boxes which most likely contain functionally important regulatory elements. Third region is 1.5 kb long and it located approximately 10 kb upstream of P1. It is noteworthy that all three regions are free from transposones, have nucleosome assembling characteristics distinguishing them from the rest of the sequences in this region and have high promoter potential.

The presence of various repetitive elements characterizes both vole and mouse 5' regions. In average, 45% of vole upstream region consists of various type of repeats, B1 family is most abundant; mouse 5'-region has 28.6% of repetitive elements. The detailed maps of the repeat localization, with designation of their orientation, size and identity to consensus sequences, were made for each species. Nucleosome assembling potential profiles were also determined for each 5' region. The maximum of the nucleosome assembling potential coincides with the site of repetitive elements localization. In contrast promoter sequences have a low nucleosome assembling potential. Taken together this characteristics with enormous amount of mobile elements in this region led to the conclusion that these SINEs might play an important role in nucleosome positioning and therefore may influence the chromatin structure of this region, and thus affect XIST transcription. The data suggest a possible involvement of described conserved sequences and repeats in Xist regulation.

Acknowledgements

This work was supported by the grant INTAS №99-00284.

Reference

1. B. D. Hendrich, R. M. Plenge, and H. F. Willard, "Identification and characterization of the human *XIST* gene promoter: implications for models of X chromosome inactivation" Nucleic Acids Res. **25:**2661–2671(1997).

HEAT SHOCK KNOWLEDGE BASE

*[1]Stepanenko I.L., [1]Levitsky V.G., [2]Schroeder H.C.

[1]Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
[2]University of Mainz, Institute for Physiological Chemistry, Mainz, Germany
e-mail: stepan@bionet.nsc.ru
*Corresponding author

Keywords: data bases, regulation transcription, ontology, heat shock

Introduction

Every cell responds to environmental, chemical, and physiological stress through a rapid and preferential increase in expression of a highly conserved group of proteins known as the heat shock proteins (HSP). The HSPs protect the cells from various stresses and can be grouped into three general classes: chaperones that act in refolding of misfolded proteins, proteases that degrade of damaged proteins and stress-specific proteins alleviating specific stress. Owing to its responsiveness to diverse forms of stress, the heat shock response has undergone widespread application in biomonitoring and environmental toxicology. The use of HSPs as biomarkers is most widespread in aquatic toxicology, so within the frames of the Project INTAS-96-1787, a computer database on heat shock genes available through the Internet was developed.

Results

The expression of HSP is primarily regulated at transcription level. The experimental data about transcription regulation of heat shock genes are collected in TRRD (Kolchanov N.A. et al., 2000). The formalised description of patterns of gene expression, regulation regions, and binding sites of transcription factors of heat shock genes are accumulated in HS-TRRD database. At present, TRRD contains the experimental data on 80 heat shock genes and 46 heat shock elements.

HEAT SHOCK GENE NETWORK section contains experimental data on gene interactions during the heat shock response. The regulation of heat shock response is represented in the form of an object-oriented gene network database GeneNet (Kolpakov F.A. et al., 1998; Kolpakov F.A., Ananko E.A., 1999). The central to heat shock response in eukaryotes is the heat shock transcription factor (HSF) that regulates the expression of HSP in response to stress. HSF recognizes the promoters of heat shock genes through DNA sequences called as heat shock elements. A key step in heat-induced transcription is activation of the heat shock transcription factor (Fig.1).

Using GeneNet format, HSKB accumulates also an information on cell cycle-specific, developmental and redox regulation of heat shock genes, heat shock response and apoptosis.

HEAT SHOCK GENES section describes organisation of HSP genes into multigene families. The heat shock-induced genes, constantly expressed genes (hsc, heat shock cognates) and heat shock transcription factor genes are also included into this section. The location of the potential nucleosome binding sites at the heat shock gene promoters and its correlation with the heat shock elements positioning were determined by computer analysis.

HEAT SHOCK ELEMENTS section contains description of transcription regulatory elements of heat shock genes. HSEs can be found in promoter region of every heat shock gene. A HSE consists of alternating, inverted repeats of the nGAAn, where n can be an arbitrary nucleotide (Tabl.1). Typical HSEs have from two to six repeats, and heat shock promoters often have more than one HSE. The level of HSP expression depends on the number of repeats, the conservation of the repeats to the nGAAn consensus , and the distance between HSE and promoter. The section contains the data of computer analysis of conformational and physicochemical DNA properties of heat shock elements (Oshchepkov D.Yu. et al., 2000)

HEAT SHOCK TRANSCRIPTION FACTORS section is devoted to description of transcription factors of heat shock gene system. Although the heat shock response phenomenon is conserved in eukaryotes, both the number and overall sequence of heat shock factors vary widely among the different species. Yeast and Drosophila appear to have a single HSF gene, whereas most vertebrates and higher plants possess by multiple HSF genes, i.e., at least three different HSF genes were isolated from human, mouse, chicken, and tomato genomes. Despite sequence divergence, all members of a HSF family have two highly conserved features: they have a helix-turn-helix DNA-binding domain and coiled-coil hydrophobic repeat domains, which mediate the trimerization of HSF. Of the HSFs coexpressed in vertebrates, HSF1 is functionally analogous to yeast and Drosophila HSF as stress-induced factor, but HSF2 activity is associated with development and differentiation. During the heat shock response, the regulation and role of HSF2 has been enigma, HSF2 is activated by the flux of non-native proteins targeted for protein degradation.

THE HEAT SHOCK PROTEIN section contains HSP classification, description of its cellular role and computer analysis of J-domains (Afonnikov D.A., 2000).

Heat shock knowledge base (HSKB) is developed in the object-oriented manner. It is based on the accumulation of complete experimental data of heat shock response and ontology of this system. Results of computer treatment of experimental data are stored in the HSKB.

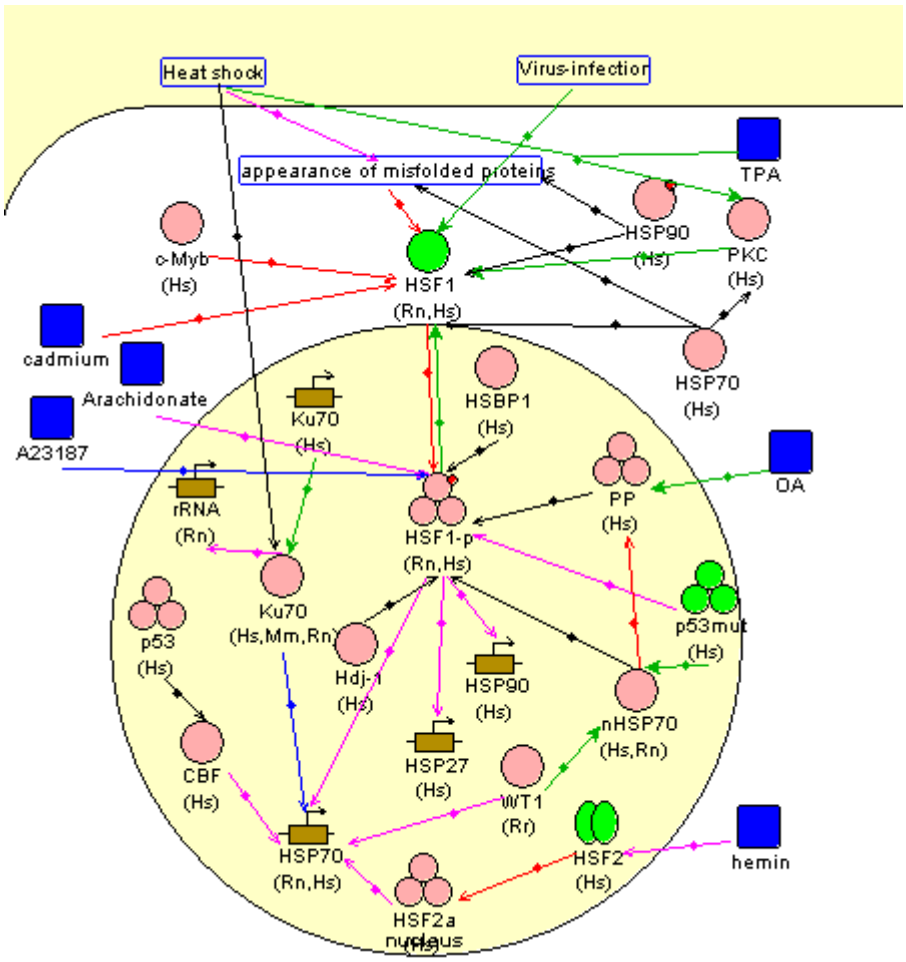


Figure 1. In the unstressed cell, HSF1 is repressed and maintained in a monomer, non-DNA-binding form through transient interactions with chaperones such as HSP70 and HSP90. Negative regulation of the DNA binding and transactivation domains involves intermolecular interactions with the central regions of HSF1 and is also influenced by constitutive phosphorylation at serine residues. Activation of HSF1 is linked to the appearance of non-native proteins and the requirement for HSP70 and HSP90 to prevent the appearance of misfolded proteins. Upon heat shock or other forms of stress, HSF1 assembles into trimmer, binds to HSE in promoters and transforms into phosphorylated form. Heat shock also represses DNA-binding ability and the synthesis of constitutive HSE-binding transcription factor (CHBF or Ku/E1BF). Regulation of inducible transcription of heat shock genes is provided through the negative feedback. Activation of HSF1 causes an increase in HSP70 transcription level followed by attenuation of response. The transcription activity of HSF1 is repressed by direct binding of HSP70, HSP90, and HSFBP1 with the subsequent dissociation of HSF1 from DNA.

Table 1. Heat shock factor binding sites.

TRRD gene ID	TRRDsite AC	heat shock element (HSE)	site positioning	EMBL site positioning
Dm:HSP83	S3784	TCcaGAAgccTCtaGAAgt TTCtaGAgacTTC	-86	X00065 :800
Dm:HSP26	S2894	ctTTCtaGAAaaTTgcaaca	-372	X12505:29
	S2895	ctCTaGAAacTTCggctctctcac	-352	X12505:51
	S2897	TTCtcGAAcTCAtgGCA	-266	X12505:124
	S2899	TTCcgGACtcTTCtaGAA	-62	X12505:338
Dm:HSP70	S112	GAAtaTTCtaGAA	-251	J01103:960
	S113	CTCgaGAAatTTC	-186	J01103:1565
	S114	TTCtcGTTgcTTCgaGAGa	-87	J01103:1664
	S115	GAAtgTTCgcGAA	-59	J01103:1692

TRRD gene ID	TRRDsite AC	heat shock element (HSE)	site positioning	EMBL site positioning
Dm:HSP23	S2891	GAAgtTTCgtGTC	-142	X03889:475
Dm:HSP27	S2902	GAAacTCCaGAA	-366	M12934:918
	S2903	GAAatGTCaaGAAgtTTCtgGTT	-351	M12934:931
	S2904	GAAagAGCcaGAAgaTGCgaGAG	-293	M12934:990
Dm:HSP22	S3448	GAAacTTC	-193	V00209:322
Sc:SSA1	S2934	TTCcaGAAcgTTC	-201	X12926:307
Sc:HSP82	S2911	GAAcaaTTCtgGTTctTTC	-229	X56886:132
	S2912	TTCgcaGAAcTTTttGAA	-206	X56886:155
	S2913	TTCtaGAAcgccgtgGAA	-176	X56886:185
Sc:SSA3	S3627	GAAagTTAtaGAAtaTTAcaGAA	-164	M36115:927
Sc:CUP1	S3747	TTCtaGAAgcaaaaaGAG	-164	K02204:1305
	S3750	TTCcgctGAAccgTTC	-135	K02204:1334
Sp:SSP1	S3917	GAAggTTCtcGTT	-350	X59987:702
Hs:HSP27	S2692	GAAggTTCcaGAt	-184	X03900:28
Hs:HSP70	S372	GAAgaCTCtgGAGagTTC	-196	M11717:77
	S374	GAAacCCCtgGAAtaTTCccGAC	-114	M11717:160
Hs:HSP90B	S3480	TTCtgGAAgaTTC	+682	J04988:1783
	S3481	TTCtgGAAgcTTC	+734	J04988:1791
Mm:HSP84	S4199	GAAcaTTCtaGTA	-441	U47056:188
Xl:HSP70	S2752	GAAagCTCgcGAAtcTTC	-248	X01102:115
	S2753	GAAacCTCgcGAAagTTC	-193	X01102:170
Mm:HSP701	S2870	GAAacCTCtgGAGagTTC	-197	M35021:380
	S2872	GAAacTGCtgGAAgaTTC	-116	M35021:461
Rn:HSP70	S3447	GAAgaTTCttGGC	-112	X75357:173
Hs:HSP70B	S3629	GAAccTTCcccgcgtTTC	-256	X13229:337
	S3630	TCCgcccGAAccTTC	-164	X13229:429
	S3632	GAAggTGCggGAAggTTC	-72	X13229:521
Hs:MDR1	S3171	GAAcaTTC	-171	X58723:903
	S3172	TCCtgGAAaTTC	-162	X58723:912
Hs:HO	S2700	GAAccTTCtgGGA	-383	X14782:158
Rn:HO	S2768	TTCtgGAAccTTCcaGAT	-289	M12129:1099
Mm:APG1	S3628	TTTccaGCTctTTCtcGAAccTGC	-165	D70845:308
Rn:ALBU	S3482	TTCtaGAAgaTCC	-438	S82890:66
Hs:SOD1	S2689	GAAtagcTTC	-1191	Z29336:2790
Cc:APOJ	S4144	TTCcaGAAagCTC	-70	CCT64CLU:451
Rn:P67	S4511	GAAagTTCgtGAA	-325	U37710:1225
Hs:AAP	S4526	GActtTTCtaGAg	-314	X12751:3386

Availability:

http://wwwmgs.bionet.nsc.ru/mgs/dbases/heatshock/

Acknowledgements

This work was supported by INTAS-96-1787.

References

1. Kolchanov N.A. et al. (2000) Transcription Regulatory Regions Database (TRRD): its status in 2000. Nucleic Acids Res., **28**, 298-301.
2. Kolpakov F.A., Ananko E.A. (1999) Interactive data input into the GeneNet database. Bioinformatics, **15**,713-714.
3. Kolpakov F.A., Ananko E.A., Kolesov G.B., Kolchanov N.A. (1998) GeneNet: a database for gene networks and its automated visualization. Bioinformatics, **14**, 529-537.
4. Afonnikov D.A. (2000) CRASP: software package for analysis of physicochemical parameters of aligned protein sequences. Proceedings of BGRS'2000 (this issue), Novosibirsk
5. Oshchepkov D.Yu., Stepanenko I.L., Afonnikov D.A. (2000) Correlation analysis of HSF binding sites conformational properties. Proceedings of BGRS'2000 (this issue), Novosibirsk

MODEL OF PHOSPHATE HIGH-AFFINITY TRANSPORT IN BACTERIA

Titov I.I., Kolchanov N.A., [1]Schröder H.C.

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
[1]Institute of Physiological Chemistry and Pathobiochemistry, Mainz, Germany
e-mail: titov@bionet.nsc.ru
*Corresponding author

Keywords: phosphate transport, permease translocation, diffusion, bacteria, ATP

Resume

Motivation:

Although phosphate metabolism in bacteria has been long studied, its certain features are yet to be explained. In particular, the nature of nonequivalence in the directions of primary inorganic phosphate transport is vague.

Results:

Described in this work is the quantitative model of primary phosphate transport in bacteria explaining why this transport is restricted to P_i uptake. Nonlinearity of and difference in the mechanisms of primary transport in the membrane and cell result in an essentially lower efficiency of P_i release compared to its uptake, virtually unobservable on the background of the secondary transport. Possible coordination of the primary and secondary transport mechanisms through regulation of accessible ATP content is discussed.

Introduction

Transport of inorganic phosphates, P_i, is an important metabolic element of many lower organisms, involved in bacterial response to changes in nutrient medium [1]. Here two mechanisms of P_i transport are distinguished. ATP controlled, the primary transport is carried out by permease system [2] through binding phosphates by one of its constituent proteins. The rest proteins cross the membrane through binding the substrate with ATP in excess. This system provides the cell with P_i, their moderate excess in the environment provided. The secondary mechanism was studied in detail in *E. coli* and A108 [3, 4]. It is responsible for both P_i cell uptake and its release due to a low-affinity chemical osmosis. In this work, we describe a model of the primary mechanism. In particular, the model explains why the primary mechanism is responsible only for the phosphate uptake.

Major simplifications of the model are connected with the membrane transport, namely:

1. the phosphate transport may be divided into membrane and cellular;
2. the chemical reactions involving phosphates on the membrane surface are reversible; and
3. the phosphate transport in membrane is controlled not by these reactions, but by diffusion of phosphate–protein complexes.

These assumptions allow us to reduce the transmembrane P_i transport to diffusive circulation of phosphate-binding proteins. Qualitatively, our model of P_i affinity transport may be represented as a conveyor. (For brevity, let us also include into P_i any intermediate compounds, such as HPO_4^-/HPO_4^{2-}; such substitution is correct provided that the phosphorus stoichiometry remains unchanged.) Specific proteins bind excess P_i outside the cell to transport them into membrane, passing to transporter proteins via a series of chemical transformations. In turn, the excess of phosphate–protein complexes leads to their drift to the other boundary to release P_i. The flow of phosphate-carrying complexes entails an opposite gradient of unloaded complexes (or their components) and, respectively, their antiport. Once activated, the P_i transport is controlled by their monocomponent diffusion to the discharge site.

Such a pronounced distinction of the membrane and cellular transport mechanisms should manifest itself not only in their parameters, but also in (essential at high P_i concentrations) deviation from ideality, for example, from Henry's law. Inside the cell, free phosphates interact as ions in an electrolyte, whereas their competition for carrier proteins is important in the membrane. Diffusion nonlinearity and heterogeneity together result in essential difference in P_i inward and outward transport rates, the effect known in the theory of diffusion in solids [5].

Theory

Transport in membrane. For the sake of simplicity, let us consider the P_i transport in membrane as a sequence of chemical transformations or one-dimensional diffusive moves. Then, let us restrict the

consideration to the case of ATP excess on the membrane and activated phosphate translocation system in it. A component of this system binds external phosphates; the other, transports them through the membrane of thickness L_m. The assumptions (b-c) that phosphate chemical transformations are reversible and fast allow relating their concentrations outside the cell and on the interphase boundary in the membrane:

$$K_m C_m = \frac{n_{m1}}{N_m - n_{m1}}, \quad (1)$$

where N_{ml} is the number of carrier proteins near the membrane surface per unit volume; K_m has the meaning of resulting thermodynamic gain from transporting external P_i at concentration C_m to these proteins; and n_{ml} is the concentration of P_i–protein complexes.

It is reasonable to assume that these complexes are neutral and stable except for the nearest vicinity of the membrane surface and to neglect cross-diffusion. Then, the excess of bound complexes in the vicinity of one membrane boundary is leveled by the diffuse flow to the opposite boundary, determined exclusively by the difference in their concentrations:

$$J_m = (n_{m2} - n_{m1}) / r_m \quad (2)$$

Here, we introduced the membrane "resistance", $r_m = L_m/D_mN_m$, for the intramembrane diffusion D_m; n_{m2} is the concentration of complexes at the inner membrane surface.

Intracellular transport. Similarly to (1), phosphate concentrations n_{m2} and n_{c1} on the inner membrane surface are determined from the equality of chemical potentials on it:

$$\mu_c(n_{c1}) = kT \ln(K_c \frac{n_{m2}}{N_{m2} - n_{m2}}) \quad (3)$$

Here, K_c characterizes the ratio of P_i solubilities in the cell and membrane at low phosphate concentrations. Such K_c specification fixes the background value wherefrom P_i chemical intracellular potential $\mu_c(n_c)$ is reckoned. Nonuniform intracellular phosphate distribution along the x-axis causes their diffuse flow:

$$J_c = -\frac{L_c}{kT}\frac{\partial \mu_c}{\partial x} \quad (4)$$

In case of the monocomponent diffusion and absence of electric field, the P_i drift is determined exclusively by the phosphate concentration gradient. Phosphates in the cell are highly ionized; therefore, let us particularize the general equation of nonequilibrium thermodynamics (4) with a simplest approximation of concentration dependence of strong electrolyte chemical potential using the Debye–Hückel theory:

$$\mu_c(n_c) = \mu_0 + kT \ln n_c + a n_c^{\frac{3}{2}} \quad (5)$$

Spatial nonequivalence of primary phosphate metabolism. The stationary phosphate flow has the equal value at any section x, resulting in equality of (2) and (4):

$$J_m = J_c \quad (6)$$

Solving the set of equations (1–6) allows the value of the stationary flow through the entire system at given external and intracellular phosphate concentrations to be found. The difference between the direct and inverse flows (under inversion of boundary conditions) is of interest. Not showing the overall cumbersome equation, let us mention that these flows are equal within a low concentration range and may differ essentially in case of considerable deviations from thermodynamic equilibrium. At the boundary phosphate concentrations C = 0 and C = 10 nM and typical cell values of other parameters the ratio of influx and efflux can be estimated as 10^2. Thus, the phosphate release within the primary mechanism is not only inferior to their uptake, but appears hardly observable, the secondary transport provided.

Discussion

The high-affinity mechanism of bacterial P_i uptake functions only in case of intermediate excess of external phosphorus. The P_i taken up by the cell is ATP-dependently utilized into polyphosphates. An excess of polyphosphates or ATP spending (by other processes as well) shifts the utilization equilibrium, and depolymerization frees the phosphates and represses the primary mechanism through decreasing the P_i concentration gradient inducing it (however, not inhibiting the affinity mechanism through saturating carrier proteins). The ATP produced through depolymerization stimulates the secondary mechanism to further phosphate uptake. In this process, the low-affinity transport creates the ATP excess on the membrane, which serves, in turn, as a substrate for the high-affinity mechanism. In other words, each of the two transport systems while operating creates the prerequisites (specifically, the ATP reservoir) for activation of its counterpart.

Our results demonstrate the inefficiency of primary transport for phosphate removal from bacteria. Possibly, the secondary mechanism appeared later for solving this problem. The tight linkage of these transport systems is also traced through a synergistic ATP content–dependent coordination (ATP, in addition, is actively involved in the intracellular biochemistry), important for adapting bacteria to the changes in environment.

Acknowledgments

The authors are grateful to Dr. B. Lorenz for helpful discussions and to G. Chirikova for translating the manuscript into English. The work was supported by the INTAS grant No. 96-1787.

References

1. Kulaev I.S. (1979) *The Biochemistry of Inorganic Polyphosphates.* New York: John Wiley & Sons, Inc, New York.
2. Bishop L., Abgayani R., Jr., Ambudkar S.V., Maloney P.C., Ames J.F.-L. (1989) *PNAS* **86**, 6953.
3. Rosenberg H., Gerdes R.G., Chegwidden K. (1977) *J. Bacteriol.* **131**, 505.
4. Van Veen H.W., Abee T., Kortstee G.J.J., Konings W.N., Zehnder A.J.B. (1994) *Biochemistry* **33**, 1766.
5. Yakobson B.I. (1987) *Physica Scripta,* **36,** 513.

METHOD FOR PREDICTION OF TRANSCRIPTION FACTOR BINDING TO REGULATORY REGION, BASED ON CAUSED BY POINT MUTATION ALTERATIONS IN THE PATTERN OF THIS REGION INTERACTION WITH NUCLEAR PROTEINS

Ponomarenko M.P., Merkulova T.I., Pomonarenko Yu.V., Orlova G.V.

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: pon@bionet.nsc.ru
*Corresponding author

Keywords: SNP's, nuclear expract, gel-shift, prediction, transcription factor, cluster-analysis

Resume

Motivation:

For several single nucleotide substitutions in regulatory regions (rSNPs) that are associated to the known clinical phenotypes, we have detected transcription factors, with the binding sites either damaged or de novo appeared due to rSNPs. These transcription factors were detected by alterations in the pattern of interaction of these regions with proteins.

Results:

We have developed the method for recognition of transcription factor (TF) binding to regulatory DNA region. This method is based on the caused by rSNPs alterations in the pattern of this region interaction with proteins. By the example of modified by rSNP genes *NTF*, K-*ras*, *TDO2*, and *pC,* we have demonstrated that the method gives the results that are in a good agreement with independent experimental data.

Availability:

The TF-site recognition tools MATRIX, http://wwwmgs.bionet.nsc.ru/mgs/programs/freq/.

Introduction

Alteration by single nucleotide position (rSNP) substitution of a pattern of this region (rDNA) interaction with nuclear proteins reflects the damage of the binding site of some transcription factor (TF-site) or its appearance de novo.These rSNPs are frequently associated with the known clinically phenotype [1-4]. We have developed a method for transcription factor (TF) prediction in accordance with rSNP-alterations in the pattern of interaction of this region with proteins. The method estimates for both strands of each rSNPs-variant of rDNA the Score-recognition of a wide variety of TF-sites. Out of this variety, those TF-sites are extracted, rSNP-altered Score of which have a significant similarity to rSNP-mediated alterations of patterns of interaction of rDNA with nuclear proteins. By this method, we have predicted transcription factors that had rSNP-damaged sites, which spoiled the functioning of the genes *NTF* [1], *pC* [2], K-*ras* [3], and *TDO2* [4]. These predictions were in acordance with these factors detected experimentally. In addition, the robustness of the prediction results to the variation of similarity scales and cluster-analysis methods is demonstrated.

Materials and Methods

The rSNP under study are given in the Table 1. Under analysis of rSNP's by the method suggested and applied to each strand of each variant of rDNA by the help of the MATRIX package ([5], http://wwwmgs.bionet.nsc.ru/mgs/programs/freq/), the maximum of the Score-recognition for each TF-site is found (Fig. 1). As a result, we get for each TF a 2xN-dimensional numerical vector (where: N is the number of rSNP-variants of rDNA, Table 2: columns I-IV). The novelty of our approach is that the set of 2xN-vectors of TFs is supplemented by the 2xN-vectors of X transcriptional factors such that rSNP-alterations in the Score in the analyzed rDNA should be assigned to the values simulating rSNP-alterations in the pattern of interaction with proteins (Table 2, lines A-D). In this case, as an estimate of statistical similarity between factors TF and X may serve the following:

$$\Delta_{TF,\,X} = D_{None,X} - D_{TF,X} - t_{\alpha;\nu} \times s.d.(D_{TF,X}) \quad (1)$$

$$\Delta_{TF,None} = -D_{TF,\,None} + t_{\alpha;\nu} \times s.d.(D_{TF,None}), \quad (2)$$

.where $D_{Q,P}$, is Euclidean distance from Q to P; $t_{\alpha;\nu}$, Student's coefficient; sd, standard deviation,

Prediction: $\{\Delta_{TF,X}>0\} \Rightarrow$ {the X factor under study is similar to the known TF factor }. (3)

As the control for robustness of prediction, obtained by the equations (1-3), we use the standard package STATISTICA: for each pair [similarity scale]+[cluster-analysis method], out of all integrity of 2xN-vectors, each X-factor under study is put in correspondence to the closest known TF-factor (Fig. 2, Table 3). This is the second novelty of our method.

Results and Discussion

Applying the method suggested (equation 3) to prediction of the TF-factor binding to rDNA, by rSNP-mediated alterations of complexes of this rDNA with the nuclear proteins, we have studied the rSNPs of the following genes: *NTF* [1], *pC* [2], K-*ras* [3], and *TDO2* [4] (Table 1). The results of predictions obtained are given in Fig. 2 and in Table 2 and Table 3. For rSNP αA of the *NTF* gene [1], by equation (3), the OCT TF-factor was predicted (Table 2), resistant to variation of similarity scales and cluster-analysis methods in STATISTICA package (Fig. 2, Table 3: 25 out of 35 combinations). This prediction coincides to results of the control testing of rSNP *NTF* αA by applying antibodies to OCT-1 [1]. For pCmt, the equation (3) has predicted the HNF-1 factor (Table 3), this fact being in coincidence with an experiment [2]. For the K-*ras*-CA gene, the GATA-factor was predicted (Table 3), which is in a good accordance to experimental evidence demonstrating competition between GATA-binding oligo-DNA to the K-*ras*-CA gene during formation of the complex analyzed with the lung nuclear [3]. For the rSNP-variants WT, M1, and M2 of the *TDO2* gene, we have predicted the factors YY1, GR, and HNF3, respectively. The presence of the first factor, YY1, was experimentally verified by the authors by using antibodies against YY1 [4].

The results obtained indicate that cluster-analysis of the Score-recognition of a wide variety of TF-sites is an adequate model for analysis of rSNP-alterations in the pattern of interaction of rDNA with nuclear proteins. Cluster-analysis of TF-sites recognition is really a novel approachn to analysis of interactions between DNA and nuclear proteins.

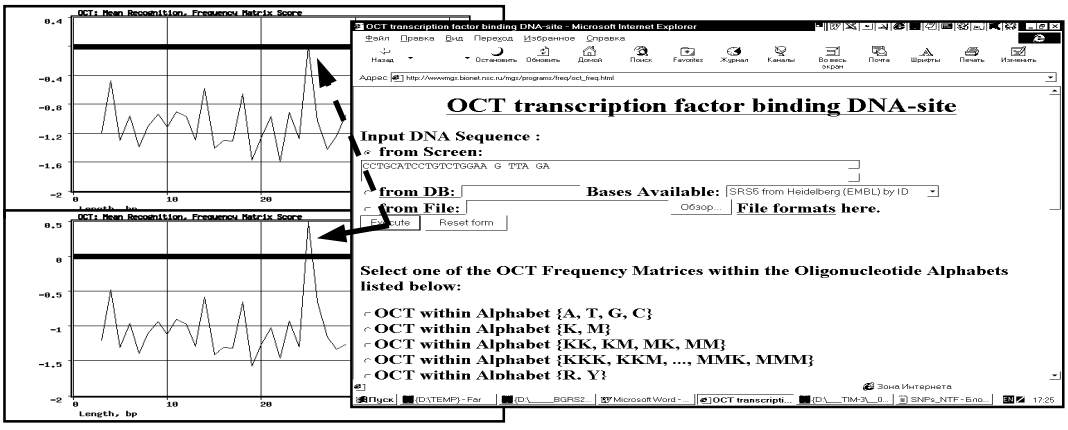


Figure 1. Recognition of the site OCT in the rSNPs *NTF* αG (continuous) and *NTF* αA (broken).

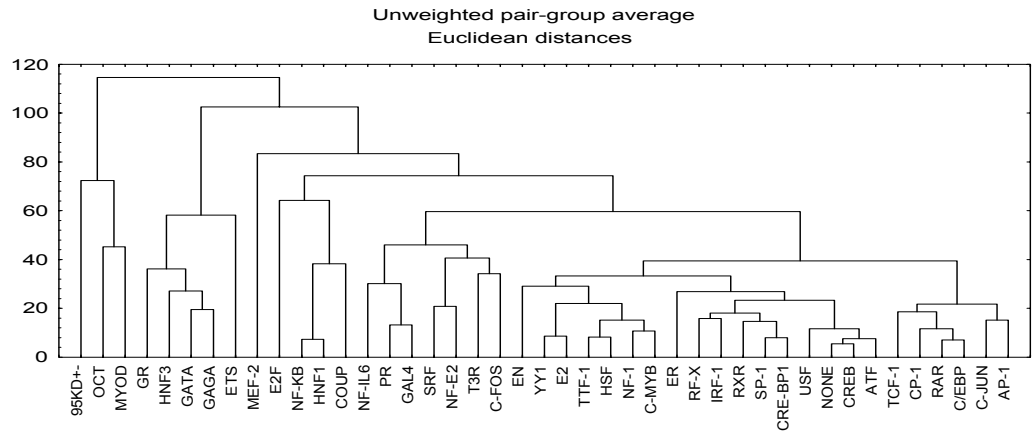


Figure 2. An example of the STATISTICA result on the data from the Table 2 (columns I-IV).

Table 1. Experimental data under study (#, multiple complexes; $, transcription activity)

	rSNP's	DNA (5'-3', positive chain)	Extract/cell	Interaction	Ref
1	*NTF* αG	CCTGCATCCTGTCTGGAA **G** TTAGA	MonoMac6	-	[1]
	NTF αA	CCTGCATCCTGTCTGGAA **A** TTAGA		+	
2	pCwt	TTTGTGGTTATGGA **T** TAACTCGAACT	HepG2	+$)	[2]
	pCmt	TTTGTGGTTATGGA **C** TAACTCGAACT		-$)	
3	K-*ras*-CA	GTGCAAGAAA **C** TCCACTT **A** TCATGAGAGCT	Lung	+	[3]
	K-*ras*-CC	GTGCAAGAAA **C** TCCACTT **C** TCATGAGAGCT		-	
	K-*ras*-GC	GTGCAAGAAA **G** TCCACTT **C** TCATGAGAGCT		-	
4	WT	CAGTTGCCAAATAATG **G** CA **G** ATAAGAATAGGGAG	Liver	++ - +-#)	[4]
	M1	CAGTTGCCAAATAATG **A** CA **G** ATAAGAATAGGGAG		- ++ -	
	M2	CAGTTGCCAAATAATG **G** CA **T** ATAAGAATAGGGAG		+- - ++#)	

Table 2. Prediction with the help of equations (1-3) of TF-sites damaged by rSNP's in the *NTF* gene.

No.	TF name	+αG	-αG	+αA	-αA	95Kd++	95Kd+-	95Kd-+	None	Best pattern
1	AP-1	-0,22	-0,30	-0,33	-0,30	-0,40	-0,58	-0,47	**0,34**	None
2	ATF	-0,38	-0,40	-0,38	-0,40	-0,49	-0,65	-0,57	**0,55**	None
3	C/EBP	-0,38	-0,22	-0,40	-0,18	-0,37	-0,70	-0,32	**0,30**	None
4	c-Fos	-0,10	-0,20	-0,10	-0,32	-0,29	-0,35	-0,52	**0,10**	None
5	c-Jun	-0,33	-0,20	-0,33	-0,33	-0,42	-0,61	-0,47	**0,35**	None
6	c-Myb	-0,22	-0,50	-0,30	-0,55	-0,56	-0,55	-0,76	**0,33**	None
7	COUP	-0,40	-0,75	-0,40	-0,75	-0,79	-0,75	-1,02	**0,13**	None
8	CP-1	-0,42	-0,27	-0,35	-0,18	-0,32	-0,66	-0,33	**0,32**	None
9	CRE-BP1	-0,53	-0,45	-0,45	-0,45	-0,58	-0,76	-0,64	**0,48**	None
10	CREB	-0,45	-0,38	-0,43	-0,40	-0,53	-0,72	-0,57	**0,55**	None
11	E2	-0,20	-0,40	-0,22	-0,43	-0,42	-0,45	-0,63	**0,32**	None
12	E2F	-0,33	-0,70	0,00	-0,70	-0,52	-0,34	-1,00	**0,02**	None
13	EN	-0,50	-0,50	-0,18	-0,45	-0,40	-0,50	-0,67	**0,34**	None
14	ER	-0,42	-0,55	-0,62	-0,45	-0,70	-0,90	-0,68	**0,36**	None
15	Ets	-0,35	0,40	-0,35	0,20	-0,29	-0,95	-0,02	-0,42	
16	GAGA	-0,52	0,19	-0,68	0,19	-0,49	-1,17	0,00	-0,30	
17	GAL4	-0,20	-0,50	-0,18	-0,18	-0,21	-0,44	-0,44	**0,20**	None
18	GATA	-0,68	0,25	-0,65	0,10	-0,54	-1,19	-0,12	-0,31	
19	GR	-0,85	0,05	-0,82	0,02	-0,69	-1,32	-0,26	-0,26	
20	HNF1	-0,65	-0,87	-0,68	-0,80	-1,04	-1,11	-1,15	-0,09	
21	HNF3	-0,77	0,22	-0,77	0,30	-0,56	-1,37	-0,01	-0,47	
22	HSF	-0,38	-0,60	-0,38	-0,60	-0,65	-0,67	-0,83	**0,34**	None
23	IRF-1	-0,65	-0,40	-0,50	-0,40	-0,59	-0,85	-0,59	**0,36**	None
24	MEF-2	-1,08	-0,25	-0,98	-0,20	-0,94	-1,49	-0,62	-0,31	
25	MyoD	0,15	-0,38	0,15	-0,38	-0,24	-0,08	-0,76	-0,19	
26	NF-1	-0,30	-0,55	-0,30	-0,50	-0,52	-0,56	-0,72	**0,38**	None
27	NF-E2	-0,18	-0,05	-0,18	-0,05	-0,16	-0,52	-0,22	-0,01	
28	NF-IL6	-0,30	-0,30	0,00	-0,10	-0,02	-0,31	-0,33	**0,06**	None
29	NF-kB	-0,65	-0,85	-0,63	-0,85	-1,04	-1,08	-1,18	-0,09	
30	OCT	0,00	-0,30	0,50	-0,15	*0,15*	**0,21**	-0,67	-0,44	**95Kd+-,** *95Kd++*
31	PR	-0,20	-0,50	-0,20	-0,05	-0,14	-0,49	-0,32	**0,12**	None
32	RAR	-0,35	-0,18	-0,45	-0,18	-0,41	-0,75	-0,31	**0,27**	None
33	RF-X	-0,60	-0,35	-0,40	-0,50	-0,59	-0,74	-0,67	**0,39**	None
34	RXR	-0,60	-0,50	-0,55	-0,55	-0,73	-0,89	-0,76	**0,33**	None
35	Sp-1	-0,50	-0,47	-0,50	-0,50	-0,65	-0,80	-0,69	**0,45**	None
36	SRF	-0,28	0,05	-0,15	0,10	-0,06	-0,59	-0,07	-0,15	
37	T3R	0,05	-0,08	0,05	-0,08	-0,07	-0,28	-0,39	-0,20	
38	TCF-1	-0,50	-0,20	-0,35	-0,30	-0,41	-0,68	-0,44	**0,34**	None
39	TTF-1	-0,35	-0,55	-0,35	-0,55	-0,59	-0,62	-0,76	**0,39**	None
40	USF	-0,35	-0,35	-0,35	-0,45	-0,51	-0,61	-0,61	**0,49**	None
41	YY1	-0,20	-0,47	-0,25	-0,47	-0,47	-0,49	-0,68	**0,32**	None
A	95Kd++	-0,42	-0,42	**1**	**1**	**1,47**	-0,69	-0,59	-1,39	*OCT*
B	95Kd+-	**0**	-0,42	**1**	-0,42	0,0	**0,78**	-1,20	-0,87	**OCT**
C	95Kd-+	-0,42	**0**	-0,42	**1**	0,0	-1,30	**0,88**	-0,87	
D	None	-0,42	-0,42	-0,42	-0,42	-0,53	-0,69	-0,59	**0,613**	**28 TF's**

Table 3. Predictions of TF-sites, damaged rSNP's, by means of the package STATISTICA.

rSNP's		D2	ED	MD	CD	W
TNF, αA	SL	**OCT**	**OCT**	MEF-2	MEF-2	**OCT**
	CL	**OCT**	**OCT**	**OCT**	**OCT**	**OCT**
	UPGA	**OCT**, MyoD	**OCT**, MyoD	**OCT**, MyoD	**OCT**	**OCT**, MyoD
	WPGA	Ets	MEF-2	**OCT**, MyoD	**OCT**	MEF-2
	UPGC	**OCT**, MyoD	MEF-2	NF-kB, HNF1	MEF-2	MEF-2
	WPGC	MEF-2	**OCT**	**OCT**	**OCT**	**OCT**
	WM	**OCT**, MyoD	**OCT**, MyoD	**OCT**, MyoD	**OCT**, Ets	**OCT**, MyoD
pCwt	SL	**HNF-1**	**HNF-1**	**HNF-1**	**HNF-1**	**HNF-1**
	CL	**HNF-1,** NF-IL6	**HNF-1**, NF-IL6	**HNF-1**	**HNF-1**	**HNF-1**, NF-IL6
	UPGA	**HNF-1**	**HNF-1**	**HNF-1**	**HNF-1,** NF-IL6	**HNF-1**
	WPGA	**HNF-1**	**HNF-1**	**HNF-1**	**HNF-1**	**HNF-1**
	UPGC	**HNF-1**	**HNF-1**	**HNF-1**	**HNF-1**	**HNF-1**
	WPGC	**HNF-1**	**HNF-1**	**HNF-1**	**HNF-1**	**HNF-1**
	WM	**HNF-1**, NF-IL6	**HNF-1**, NF-IL6	**HNF-1**	**HNF-1**	**HNF-1**, NF-IL6
K-*ras*-CA	SL	**GATA**	**GATA**	**GATA**	**GATA**	**GATA**
	CL	**GATA**	**GATA**	**GATA**	**GATA**	**GATA**
	UPGA	**GATA**	**GATA**	**GATA**	**GATA**	**GATA**
	WPGA	**GATA**	**GATA**	**GATA**	**GATA**	**GATA**
	UPGC	**GATA**	**GATA**	**GATA**	**GATA**	**GATA**
	WPGC	**GATA**	**GATA**	**GATA**	**GATA**	**GATA**
	WM	**GATA**	**GATA**	**GATA**	**GATA**	**GATA**
TDO2	SL	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**
		M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**
		M2+-≈GAGA	M2+-≈GAGA	M2+-≈OCT	M2+-≈Sp1	M2+-≈GAGA
	CL	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**
		M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**
		M2+-≈**HNF3**	M2+-≈**HNF3**	M2+-≈**HNF3**	M2+-≈TCF-1	M2+-≈**HNF3**
	UPGA	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**
		M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**
		M2+-≈TCF-1	M2+-≈TCF-1	M2+-≈**HNF3**	M2+-≈SRF	M2+-≈TCF-1
	WPGA	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**
		M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**
		M2+-≈**HNF3**	M2+-≈TCF-1	M2+-≈**HNF3**	M2+-≈SRF	M2+-≈TCF-1
	UPGC	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**
		M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**
		M2+-≈**HNF3**	M2+-≈GAGA	M2+-≈**HNF3**	M2+-≈SRF	M2+-≈OCT
	WPGC	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**
		M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**
		M2+-≈TCF-1	M2+-≈TCF-1	M2+-≈**HNF3**	M2+-≈SRF	M2+-≈TCF-1
	WM	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**	WT+-≈**YY1**
		M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**	M1+-≈**GR**
		M2+-≈TCF-1	M2+-≈TCF-1	M2+-≈SRF	M2+-≈TCF-1	M2+-≈TCF-1

SL, Single Linkage; CL, Complete linkage; UPGA, Unweighted Pair-Group Average; WPGA, Weighted Pair-Group Average; UPGC, Unweighted Pair-Group Centroid; WPGC, Weighted Pair-Group Centroid; WM, Ward's Method. D2, Squared Euclidean Distance; ED, Euclidean Distance; MD, Manhattan Distance; CD, Chebyshev's Distance; W, Power. Predictions by the equation (3) are marked by bold, the results given by STATISTICA (Fig. 2) are in a frame.

This work was supported by the grants RFBR 98-07-90126, 00-04-49548, and INTAS 96-1787.

References

1. Knight,JC, Udalova,I, et al. (1999) *Nuture Genetics*, **22**, 145-150
2. Spek,CA, Lannoy,VJ, et al. (1998) *J.Biol. Chem.*, **273**. 10168-10173
3. Levashova Z.B., Kaledin V.I., Ponomarenko M.P., BGRS-2000 (this issue)
4. Vasiliev G.V., Merkulov, V.M., Kobzev V.F., et al. (1999) *FEBS Lett.,* **462**, 85-88.
5. Ponomarenko M.P., et al. (1999) *Bioinformatics*, **15**, 631-643.

rSNP_GUIDE: A DATABASE DOCUMENTING INFLUENCE OF SUBSTITUTIONS IN REGULATORY GENE REGIONS ONTO THEIR INTERACTION WITH NUCLEAR PROTEINS AND PREDICTING PROTEIN BINDING SITES, DAMAGED OR APPEARED DE NOVO DUE TO THESE SUBSTITUTIONS

*Ponomarenko J.V., Orlova G.V., Ponomarenko M.P., Lavryushev S.V., Merkulova T.I.

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: jpon@bionet.nsc.ru
*Corresponding author

Keywords: SNP, clinical phenotype, regulatory DNA, transcription factor, prediction, guide-base

Resume

Motivation:

Modern databases contain descriptions of more than 150 point mutations in regulatory gene regions, these mutations are being associated with clinical phenotypes appeared due to gene expression disorders.

Results:

We have developed a database rSNP_Guide accumulating the data on the influence of substitutions with the known clinical phenotypes onto interaction of regulatory DNA with the proteins. The novelty of our database is an HTML-guide including prescriptions for prediction of transcription factors that have the sites damaged by single nucleotide substitutions.

Availability:

The database rSNPdb, URL=http://wwwmgs.bionet.nsc.ru/systems/rsnpdb/.

Introduction

Point nucleotide substitutions (Single Nucleotide Polymorphisms, SNP) are the reasons of one of the most widespread genome polymorphism. During the recent years, the tremendous growth of experimental data on SNP is caused by the relatedness of these data to such clinical phenotypes as disease susceptibility and drug resistance/sensitiveness. Experimental evidence on SNP is stored in such databases as HGMD [1], dbSNP [2], HGBASE [3], ALFRED [4], and OMIM [5]. According by the influence on the clinical phenotype, most of SNPs may be subdivided into three groups: 1) cSNP, modifying codons in the genes of proteins; 2) sSNP, modifying splicing sites; 3) rSNP, modifying gene expression pattern. The HGMD database [1] accumulates more than 10000 cSNP, about 1500 sSNP, and 168 rSNP. As can be seen, the data on cSNP and sSNP are considerably prevailing over the data on rSNP. The same observation is characteristic for the other types of databases on SNP [2-5]. This fact may be explained by rather more developed experimental methods that reveal cSNP and sSNP. Besides, it is more likely to detect the reason of appearance of cSNP or sSNP with the known clinical phenotypes via destruction of protein structure and function. Thus, rSNPs modifying gene expression regulation stay less studied. On the other hand, the databases TRRD [6], COMPEL [7], TRANSFAC [8], RegulonDB [9], EpoDB [10], and ACTIVITY [11] cite above 2500 papers devoted to artificial point mutations in regulatory gene regions, such that alteration in DNA-protein interaction enabled to determine in an experiment the natural transcription factor binding site (TF-site). Moreover, artificial damaging of TF-sites were characterized by clinically determined phenotypes, thus, demonstrating the vital importance of these sites. Basing on collection of references stored in the above databases, we have developed a rSNP_Guide database on the influence of point substitutions in regulatory DNA region, with clinical phenotypes, on the pattern of these regions' interaction with proteins. The novelty of the rSNP_Guide is a supplementary database rSNP_Tools storing the prescriptions, how to implement the MATRIX database designed for TF-sites recognition [12] for analysis of the rSNP with the aim to predict TF-sites, the damage or appearance *de novo* of which may explain the clinical phenotypes of these rSNPs. This novelty of our approach is addressed to analysis of novel rSNPs that could be analyzed by analogy to the already known ones. If an analog for the novel rSNP will be found among the already known rSNPs, then the example of MATRIX [12] implementation to the known rSNP may serve as a prescription for analysis of the novel rSNP. That is why we have named the database rSNP_Guide as a "guide-base".

Materials and Methods

rSNP_Guide has three main sub-databases: rSNP_DB accumulating experimental data; SYSTEM accumulating the details of experimental design; and rSNP_Tools accumulating the examples of successful implementation of the MATRIX system on recognition of the TF-sites [12] to analysis of the known rSNP. In Fig.1a, there is an example of a document of the sub-database rSNP_DB for the point substitution T=>C in -14 position of the core promoter of the human *NTF* gene [13] (the fields **MI**, **MN**, the name of a rSNP; **DS**, clinical phenotype; **OS, OG, GR, FF, MT**, description of rSNP; **AT, AU**, **SC, SA**, brief description of the experiment on the influence of the rSNP on the pattern of interaction of regulatory region with the proteins; **FT, DR, WW**, Web-resources with supplementary information about this rSNP). The sub-database rSNP_DB stores the experimental data on rSNP presented in a format "computable" for their further computer-assisted analysis. In Fig. 1в, one can see an example of the document SYSTEM (**EP**, the goal indicated by the authors of an experiment; **EC, EM, AM, AC**, experimental conditions; **EE**, conclusions made by the authors of an experiment). SYSTEM restricts the interpretation of the rSNP analysis results to the conditions of an experiment designed for analysis of these data.

The central part of Fig. 2 gives an example of the document from the sub-database rSNP_Tools, containing implementation of the MATRIX system on recognition of the TF-sites [12] to analysis of the rSNP. The novelty of the database rSNP_Guide is the documenting of successful examples of implementation of the MATRIX system [12] to prediction of the TF-site, damaging or *de novo* appearance of which may explain the clinical phenotype of the rSNP. In case by means of the SRS-system-mediated query [13], an analog for a novel rSNP with the already known ones is found within the sub-databases rSNP_DB and SYSTEM, then rSNP_Tools gives a prescription, how to implement the MATRIX system [12] to this known rSNP. This, in its turn, may serve as a receipt for analysis of the novel rSNP. That is why we have referred the rSNP_Guide database to the "guide-base" type.

Results and Discussion

Since SRS-formats [13] and the structure of the documents stored in the sub-databases rSNP_DB and SYSTEM (Fig. 1) are analogous to the previously developed by us databases ACTIVITY [10], MATRIX [12], B-DNA-FEATURES [15], SELEX_DB [16], we focus an attention only on the format of the database rSNP_Tools (Fig. 2), which is the novelty of the present work. An example of the document rSNP_Tools for the rSNP represented in the Fig. 1 is demonstrated in the central part of the Fig. 2. This is a table of hyper-links to the Web-resources used during analysis of the rSNP by our MATRIX system developed for recognition of the TF-sites [12]. All hyper-links are grouped into five sub-tables, according to the number of steps needed for subsequent rSNP analysis.

Step #1, to recognize all TF-sites in regulatory DNA region containing the rSNP under study. For each TF-site, it is necessary to click its name (arrow 1) and to activate a Web-tool for its recognition. Then by step-wise filling the window "Input DNA Sequence From Screen" with the DNA sequence for each strand variant of the rSNP in this region (arrows 2 and 5), by clicking the option "Execute", one can get the recognition Score of the TF-site considered (arrows 3 and 6) and document the maximal value of this Score (arrows 4 and 7). As the result, each TF-site will be characterized by the 2xN-dimensional numerical vector (here: N is the number of variants of the rSNP).

Step #2, to construct the 2xN-dimensional vectors simulating the rSNP-alterations of the pattern of interaction of the DNA under study with the proteins. In our example, we have simulated 4 variants of the TF-site of interest: (1) WT++, the site is present at both strands of the natural DNA and absent in the mutant DNA (palindrome); (2) WT+-, the site is present only at (+)-strand of the natural DNA; (3) WT-+, the site is present only at (-)-strand of the natural DNA; (4) No, neither site is present. Besides, "+1" simulates the presence of the site; the mean Score of all the negative Scores documented at the Step #1 simulates the absence of the site; "0" simulates the case like "no ideas". The examples of complete sets of 2xN-dimensional vectors, one can find in our recent papers [17, 18].

Step #3, to calculate the Euclidean distances, d(X-Z), spreading from each simulated 2xN-dimensional vector X to the other such vector Z (the known TF or the X-site that should be detected).

Step #4, to detect significant by the Student t-criterion hypotheses { d(X-FT) < d(X-No) } supporting the evidence that rSNP is more likely caused by the presence of the known TF-site than by the "absence of any site". The result of this step is prediction with the aim of the MATRIX system [12] of that very particular TF-site, the damage or appearance de novo of which leads to the given clinical phenotype of the rSNP.

Step #5, to test the robustness of the statistical prediction (by the Student t-criterion). For this purpose, all 2xN-dimensional vectors obatined at the Steps #1 and #2 should be introduced into the standard package STATISTICA. For each pair [Similarity Score]+[method of cluster-analysis], it is necessary to find for every X under study the nearest according to similarity known TF-site (arrow 8). The control of statistical predictions by means of cluster-analysis is one more our novelty.

Since the interpretation of the clinical phenotypes of the rSNPs via damaging/appearing of the Tf-sites looks more complicate in comparison to cSNP и and sSNP interpretations made in accordance with protein damages, the database rSNP_Guide is addressed to overcome these difficulties while planning experiments aimed to detection of the novel rSNP.

The present work was supported by the Russian Foundation for basic research (grants Nos. 98-07-90126, 00-04-49548), and INTAS 96-1787.

References

6. Krawczak M., Ball E.V., et al. (2000) *Hum. Mutat.*, 2000; **15**, 45-51.
7. Smigielski E.M., Sirotkin K., et al. (2000) *Nucl. Acids Res.*, **28**, 352-355.
8. Brookes A.J., Lehvaslaiho H., et al. (2000) *Nucl. Acids Res.*, **28**, 356-360
9. Cheung K., Osier M.V., et al. (2000) *Nucl. Acids Res.*, **28**, 361-363.
10. McKusick, V.A.: Mendelian Inheritance in Man. Catalogs of Human Genes and Genetic Disorders. Baltimore: Johns Hopkins University Press, 1998 (12th edition).
11. Kolchanov N.A., Podkolodnaya O.A., et al. (2000) *Nucl. Acids Res.*, **28**, 298-301.
12. Kel-Margoulis O.V., Romashchenko A.G., et al. (2000) *Nucl. Acids Res.*, **28**, 311-315.
13. Wingender E., Chen X., Hehl R., et al. (2000) *Nucl. Acids Res.*, **28**, 316-319.
14. Salgado H., Santos-Zavaleta A., Gama-Castro S., et al. (2000) *Nucl. Acids Res.*, **28**, 65-67.
15. Stoeckert C. Jr., Salas F., Brunk B., Overton G.C. (1999) *Nucleic Acids Res.,* **27**, 200-203.
16. Ponomarenko M., Ponomarenko J., Frolov A., et al. (1999) *Bioinformatics*, **15,** 687-703.
17. Ponomarenko M., et al. (1999) *Bioinformatics*, **15**, 631-643.
18. Spek,CA, Lannoy,VJ, et al. (1998) *J.Biol. Chem.*, **273**. 10168-10173
19. Etzold T. and Argos P. (1993) *Comput. Applic. Biosci.*, **9,** 49-57.
20. Ponomarenko J., Ponomarenko M., Frolov A., et al (1999) *Bioinformatics*, **15,** 654-668.
21. Ponomarenko J., Orlova G., Ponomarenko M., et al. (2000) *Nucl. Acids Res.*, **28**, 205-208.
22. Ponomarenko M., Merkulova T., Ponomarenko J., et al. (2000) BGRS-2000 (this essue).
23. Levashova Z.B., Kaledin V.I., Ponomarenko M.P., , et al. (2000) BGRS-2000 (this essue)/

a

```
MI   rSNP00J0001
MN   The -14T/C rSNP  in the human protein C gene
MN   promoter reduces basal expression in HepG2
DR   SCIENTIST: SCI00002
YY
DS   type I protein C deficiency
DR   OMIM: http://…/dispmim?176860
YY
OS   human
OG   protein C gene
GR   core promoter
FF   HNF1/HNF6-binding site
MT   -14T/C, relative to trascription start
DR   HGMD: http://…/120317.html
YY
DR   rSNP_BIB: RFrSNP0J0001
WW   FIGURE: http://…/snp00j0001_fig4.html
YY
AT   basal transcription activity of CAT gene-reporter
AU   %, relative to wild-type promoter
SC   wt+, wild type, (+) chain
     TTTGTGGTTATGGA T TAACTCGAACT
SA   100
SC   mt+, mutant, (+) chain
     TTTGTGGTTATGGA C TAACTCGAACT
SA   10
DR   SYSTEM: T0rSNP00J0001
YY
TF   rSNP-damage of TF-site HNF-1 is predictable
DR   rSNP_Tools: P0rSNP0J0001
//
```

b

```
MI   T0sSNP0J0001
MN   The -14T/C rSNP in the human protein C gene
MN   promoter reduces basal expression in HepG2l
DR   SCIENTIST: SCI00002
YY
DR   rSNP_BIB: RFrSNP0J0001
DR   rSNP_DB:: rSNP00J0001
YY
EP   the aim is to study of the influence -14T/C
EP   mutation in the human protein C gene
EP   promoter on basal transcriptional activity
YY
EC   in vivo
EM   reporter construct = plasmid pCCAT493
EM   containing -396/+122 human promoter
EM   protein C region plus CAT gene the human
EM   hepatoma cell line HepG2
AM   CAT assay
AC   control transcription template contains
AC   pCwtCAT reporter construct
YY
EE   conclusion is -14T/C substitution severly
EE   reduced basal activity of the human protein C
EE   gene promoter
YY
TF   rSNP-damage of TF-site HNF-1 is predictable
DR   rSNP_Tools: P0rSNP0J0001
YY
TF   TF-site HNF-1 damaged by this -14T/C-rSNP
DR   SYSTEM: T0rSNP00J0001a
//
```

Figure 1. The examples of the SRS-documents from (a) the database rSNP_DB on experimental data on the influence of the nucleotide substitutions related to disorders with clinical phenotypes onto interaction of the regulatory DNA with the mixture of proteins; (b) the database SYSTEM, storing the experiments of the experiment conducted.

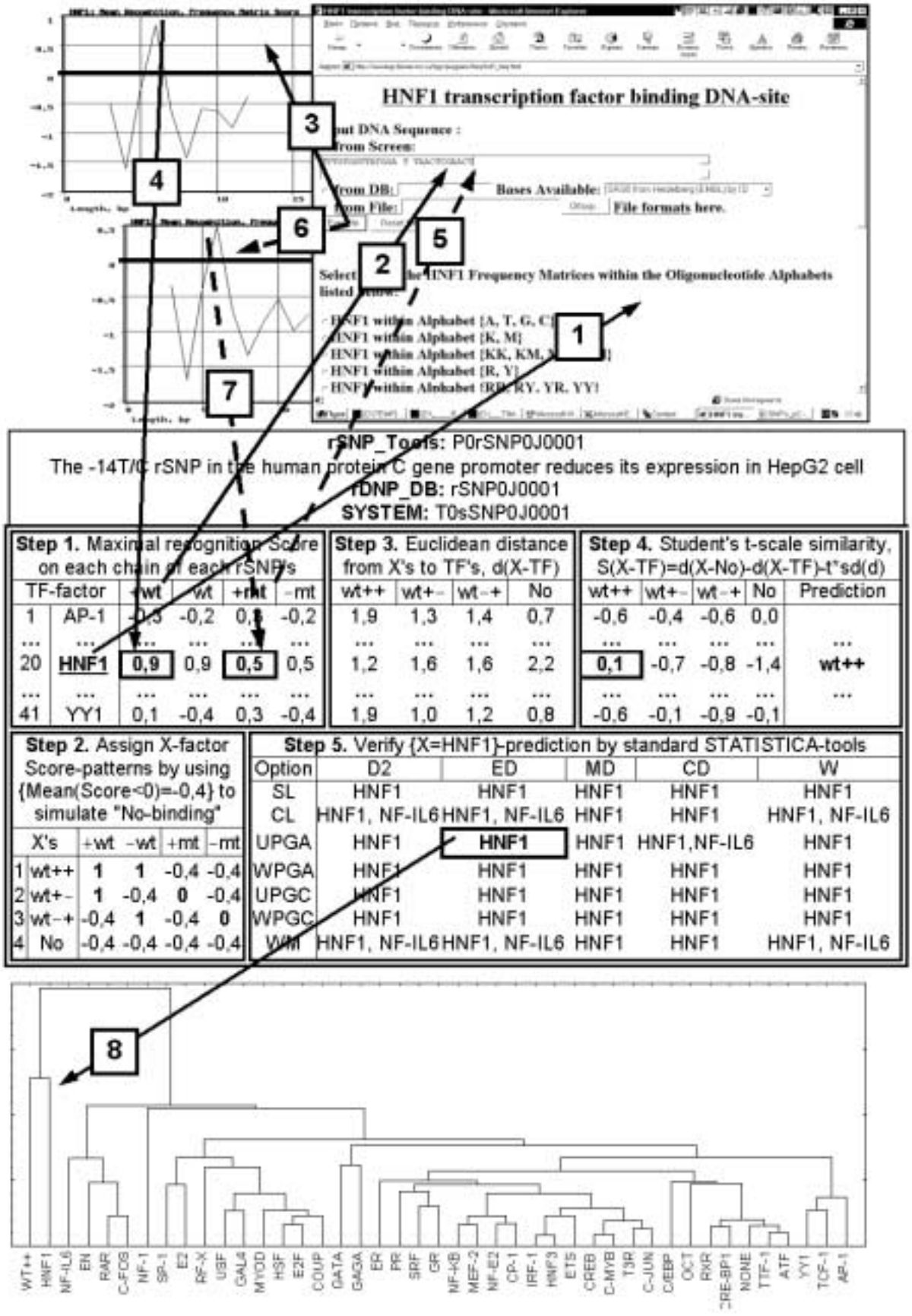


Figure 2. Example of HTML-entry of the guide-base rSNP_Guide (arrow, hyperlink/output).

COMPUTATIONAL BIOCHEMISTRY OF HEPATOCYTE.

[1,2]Demin O.V., [3]Kholodenko B.N., [4]Riznichenko G.Yu., [4]Lebedeva G.

[1]A.N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119899, Vorobjovy Gory, Moscow, Russia
[2]M.M. Shemyakin and Yu.A. Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia
[3]Department of Pathology, Anatomy and Cell Biology, Thomas Jefferson University, 1020 Locust St., 271, JAH, Philadelphia, PA 19107, USA
[4]Department of Biophysics, Faculty of Biology, Moscow State University, Vorobjovy Gory, Moscow, Russia

In this presentation we propose to develop an approach which makes «alive» all available kinetic/structural data on signal transduction and energy metabolism of hepatocyte. This implies mathematical reconstruction of following biochemical systems:

1. signal transduction pathways initiated by different growth factors such as epidermal growth factor (EGF), hepatocyte growth factor (HGF)
2. Catalytic cycle of small G-protein Ras
3. MAP kinases cascades
4. Phosphoinositol metabolism
5. Ca^{2+} signaling network
6. Glycolysis
7. Glyconeogenesis
8. Energy metabolism of mitochondria respiring on succinate, TMPD/ascorbate, glutamate/malate, pyruvate, fatty acids. This implies development of kinetic models of following subsystems of mitochondrial energy metabolism:
   - respiratory chain
   - Krebs cycle
   - ATP-synthesis
   - Connection to glycolysis (pyruvate dehydrogenase)
9. Intra- and extra- mitochondrial ammonia metabolism with connections to Krebs cycle and amino acids synthesis pathways
10. Superoxide production in mitochondria
11. Antioxidant system of mitochondria

To this aim we made use of so-called computational biochemistry, i.e., we have implemented the available kinetic knowledge concerning the components of the systems and then used a computer model to calculate the implications of that knowledge. The results of this methodology are really only dependent on the molecular kinetic information that is put in at the beginning. Hence its conclusions are just as good or as bad as the biochemistry that goes in and this is why we refer to it as computational biochemistry.

Development of kinetic models of the metabolic systems and signaling networks is accomplished in several steps. The first one is to establish kinetic scheme of the system, i.e. to find out all intermediates and all chemical reactions participating in the metabolic pathway/signaling network. The next step consists in collecting all available kinetic information concerning rates of individual reactions. This implies search and statistical processing experimentally measured data on kinetic and/or rate constants (K_m, K_{eq}, K_i, V_{max}, k_i, k_{-i} and so on). To integrate individual reaction mechanisms into the dynamic behavior of a signaling network or metabolic pathway, we convert the kinetic scheme into a set of differential equations, known as chemical kinetics equations. The resulted system of the ordinary differential equations is solved numerically. As a software we use DBSolve and SCAMP packages. All models developed in such a way is tested against experimental data.

We applied this approach to develop kinetic model of mitochondria respiring on succinate (Demin et al, 1998 Biochemistry (Moscow)). This model enabled us to reveal differences in regulatory properties of mitochondrial energy metabolism of hepatocytes extracted from control and ethanol-fed rats (Marcinkeviciute, 2000 Biochem. J.) as well as to predict how ability of mitochondria to produce superoxide depended on its energy state (Demin et al, 1998 Mol. Cell. Biochem.).

To gain a better understanding of the regulation of hepatocyte responses to EGF, we created a computational model of EGFR signaling pathway based on a detailed scheme of protein-protein interactions (Kholodenko et al, 1999 J. Biol. Chem.).

The work was supported by the INTAS grant.

References.

1. Demin, O.V., Kholodenko, B.N., Skulachev, V.P. A model of superoxide generation in the complex III of the electron transport chain. (1998) Mol. Cell. Biochem. 184, 21-33.

2. Demin, O.V., Westerhoff, H.V., Kholodenko, B.N. Mathematical modeling of superoxide generation by bc1 complex of mitochondria (1998) Biochemistry (Moscow) 63, N6, 37-53.
3. Kholodenko, B.N., Demin, O.V., Moehren, G., Hoek, J.B. Quantification of short term signaling by the epidermal growth factor receptor (1999) J. Biol. Chem. 274, 30169-30181
4. Marcinkeviciute, A., Mildaziene, V., Crumm, S., Demin, O.V., Hoek, J.B., Kholodenko, B.N. Kinetics and control of oxidative phosphorylation in rat liver mitochondria after chronic ethanol feeding. (2000) Biochem. J. (in press).

INTERACTION OF HUMAN MILK LACTOFERRIN WITH DNA IN CONTEXT OF IT'S POLYFUNCTIONAL BIOLOGICAL FUNCTIONS

Kanyshkova T.G., Babina S.E., Gal'vita A.V., Nevinsky G.A.

Novosibirsk Institute of Bioorganic Chemistry, Novosibirsk, Russia
e-mail: kanyshk@niboch.nsc.ru
*Corresponding author

Keywords: human lactoferrin, nuclease, allosteric activation

Resume

Motivation:

Many different unique functions have been attributed to lactoferrin (LF), including protection from microbial infections; immunomodulation; cell growth regulation; DNA and RNA binding, and transport into the nucleus, where LF binds specific DNA sequences and activates transcription. The physiological role of LF is still unclear, but it has been suggested that LF-DNA interactions could be the center point of numerous biological functions of this extremely polyfunctional protein.

Results:

Here we present evidence that LF is the major human milk nuclease. DNase site of this protein is localized in the N-terminal domain of LF also called lactoferricin. DNA-hydrolyzing activity of LF is activated with ATP or NAD. Taking in to account that the ATP-binding site is localized in the C-lobe of LF, and not co-interacts with protein nuclease site we proposed an allosteric mechanism of DNase function activation. Our results taken together with previously reported in the literature suggest the existence of a new transcriptional activator possessing nuclease activity, which action may depend on it's conformation state.

Introduction

LF, a major iron-binding glycoprotein of human milk, has a broad distribution in the body fluids. This protein shares a high degree of structural homology with other members of the transferrin family and possess bilobal structure with ~ 40% conservation between the N- and C-terminal lobes.

A number of biological functions, including well known antibacterial and antiviral activities, have been attributed to LF. In particular, a role in the regulation of myelopoiesis has been described by several different groups of investigators. LF was also shown to regulate granulopoiesis, antibody-dependent cytotoxicity, cytokine production, and growth of some cells *in vitro*. In addition to the above mentioned functions of LF, it is a potent activator of natural killer cells, which may play a role in antitumor defence. Thus, LF is known as an extremely polyfunctional protein, many functions of which appear to be at least in part independent of its iron-binding activity. The nature of such functions remains to be unclear yet. On the other hand, some of these functions may be the result of LF binding to specific DNA sequences. Recent research shows that LF enters the cell from the surrounding environment and is transported into the nucleus where it binds DNA. Specific DNA sequences that can confer lactoferrin-induced gene transcription of a reporter gene have now been identified. But the mechanism of such transcription regulation is not investigated.

We have previously reported that LF possess two anti-cooperative DNA-binding sites located in N-domain of protein molecule [1] and ATP-binding site located in C-domain of LF [2, 3]. The later one being bind ATP leads to dissociation of LF tetrameric form presented in normal to monomers.

Here we present evidence that lactoferrin is the unique DNase of human milk activated by ATP or NAD. Moreover, the influence of ATP on the LF interaction with specific and nonspecific DNA we observed, it is reasonable to suggest that such allosteric properties of LF may be very important in the context of its polyfunctional biological functions. Taking in to account the ability of LF to enter into the nucleus and it's influence on gene transcription it seems reasonable to believe that catalytic activity of LF may be a key mechanism of multiply regulatory properties of this protein.

Methods and algorithms

Electrophoretically and immunologically homogeneous LF was obtained by sequential chromatography of human milk proteins on DEAE-cellulose, heparin-Sepharose and anti-human LF-Sepharose column [1]. Limited tryptic cleavage of LF was performed using a trypsin concentration of 0.1-0.5% (w/w of substrate) in 0.1 M Tris-

HCl, pH 8.2, 25 mM $CaCl_2$ at 37°C for 4h. DNase activity of LF was analyzed using plasmid DNA (pBR322) and 5'-[^{32}P]oligonucleotides under optimal conditions (20 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1 mM EDTA, 37° C). *In situ* DNase activity was detected in a 12,5% SDS-PAGE gel containing 30 □g/ml calf thymus DNA as described previously [4]. To reveal the products of the DNA hydrolysis, the gel was stained with ethidium bromide; the protein were revealed by Coomassie blue staining.

Results and discussion

As we have shown previously, there is no any remarkable homology between LF and other transcriptional factors, it seems reasonable to suppose that LF may be an unique transcriptional regulator [3]. In terms of functional activity LF-DNA interactions could be the center point of numerous biological functions of lactoferrin. It's N-terminal fragment contains a polycationic amino acid tract formed two sequences; the former consists mainly of a loop of 18 amino acid residues (20-38) and the latter located primarily in N-terminus (amino acid residues 1-5). Both clusters of LF are involved in LF binding with polyanions such as heparin and lipopolysaccharides. It seems likely that it is the N-lobe fragment that is responsible for cell receptor recognition by LF and especially for DNA binding. But important as this only DNA-binding phenomenon it does not cover the problem of the LF influence to growth of some cells on the whole.

DNase activity of LF

Two further methods provided direct evidence that the LF possessed DNA-binding sites and enzymatic activity. As was shown previously [1, 3], affinity labeling of LF by chemically reactive analogs of oligonucleotides leads to modification of the protein (N-terminus of the protein). Second, after separation of the LF by SDS-PAGE in a gel containing DNA *an in-gel assay* showed DNase activity in the protein bands corresponding to LF and it's N-domain obtained using trypsin (Fig. 1).

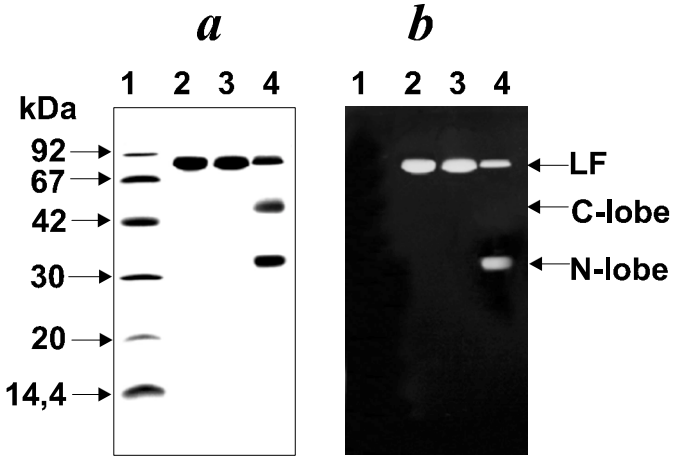


Figure 1. *In situ* DNase activity of LF and its N-lobe fragment. Line *1* – molecular mass markers, *2* – apo-LF, *3* – Fe-LF, *4* – Fe-LF digested with trypsin. To reveal activity the gel was stained with ethidium bromide (*b*), proteins were visualized with Coomassie R-250 (*a*).

As mention above, LF possesses two anti-cooperative DNA-binding sites. We demonstrated that only one of this sites is responsible for LF hydrolytic activity, namely the first site with high affinity for DNA (*see* Table 1).

Table 1. Characteristic parameters of LF-oligonucleotide binding and hydrolysis.

Substrate	K_d^1, μM	K_d^2, μM	K_m, μM	k_{cat} •10^3, min^{-1}
$d(pA)_{10}$	0.47	21.5	4.18	6.36
$d(pT)_{10}$	0.50	25.0	5.63	42.8
$d(pC)_{10}$	0.60	25.4	4.45	28.1
ON2	0.08	>100	3.69	8.2
ON3	0.22	26.1	5.29	42.2

Activation of LF DNase activity with ATP and NAD

As previously reported, LF has a strong homology with various known ATP- and NAD-binding proteins. The ATP-binding site is localized in the C-terminal domain of LF (G_{475}-M_{604}). Binding of ATP leads to dissociation of LF oligomeric forms which accompanied by a change of it's interaction with and some proteins [2]. As shown in Fig. 2, *a*, the increase of ATP concentration leads to a decrease of the affinity of LF to oligonucleotides (both specific and nonspecific). On the other hand, the increase of ATP concentration results in activation of LF in DNA (Fig. 2, *b*). There are two possible explanations for such phenomena. Firstly, the monomeric form of LF (produced by ATP binding) may be more active in DNA hydrolysis. However, a decrease of the affinity of LF for DNA in the presence of ATP (or NAD) can speak in favor of a considerable conformational change in LF polypeptide chain, especially N-domain, which responsible for protein-DNA interaction.

Some interesting features are apparent in the LF structure. The N-lobe is wide open for it's Fe-unsaturated form (apo-LF) but has closed structure for Fe-LF. Remarkably, the C-lobe cleft is closed in the case of both apo-LF and Fe-LF. The "open-close" conformation of LF N-domain is in ligand control, the main but not only one been iron ions.

We proposed that ATP (or NAD) binding by C-lobe of LF leads to conformation change in N-lobe. If this is so, the increase of DNase activity may be explain the more optimal protein conformation for DNA hydrolysis but not for DNA binding. Finally, such allosteric activation may play a key role in cell nucleus in the aspects of cell growth regulation function of LF.

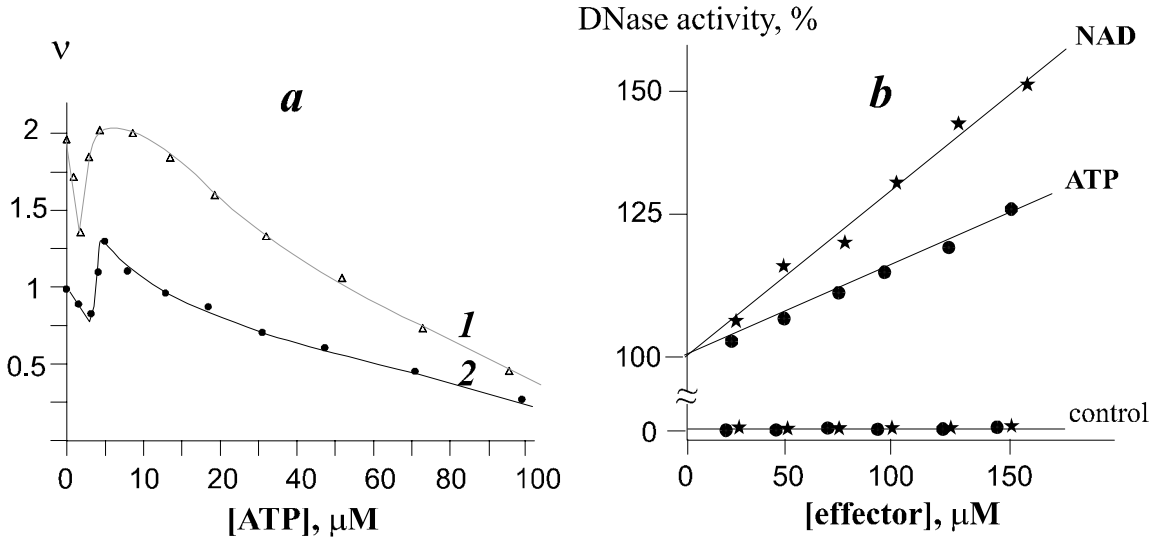


Figure 2. An influence of ATP (and NAD) on: (a) – the affinity of LF for d(pT)10 in the case of complete protein saturation by ON (1) and a half saturation (2); (b) - DNase activity of LF.

Acknowledgments

The research was made possible in part by INTAS No. 99-1645 and grant (99-04-49260) from the Russian Fund of Basic Research and a grant for young scientists from the Siberian Division of Russian Academy of Sciences.

References

1. Kanyshkova T.G., Semenov D.V., Buneva V.N., Nevinsky G.A. (1999) Human milk lactoferrin binds two DNA molecules with different affinities. *FEBS Lett.*, **451**, 235-237.
2. Semenov D.V., Kanyshkova T.G., Buneva V.N., Nevinsky G.A. (1998) Human milk lactoferrin binds and dissociates into monomers. Bioch. *Mol. Biol. Internat.*, **47**, 177-184.
3. Kanyshkova T.G., Semenov D.V., Buneva V.N., Nevinsky G.A. (1998) Lactoferrin is a new transcriptional factor having many different unique biological functions. *Proceedings of the first International Conference on Bioinformatics of Genome Regulation and Structure, Novosibirsk - Altai mountains, Russia, August 24-31*, **2**, 405-407.
4. Kanyshkova T.G., Semenov D.V., Khlimankov D.Yu., Buneva V.N., Nevinsky G.A. (1997) DNA -hydrolyzing activity of the light chain of IgG antibodies from milk of healthy human mothers. *FEBS Lett.*, **416**, 23-26.

BGRS 2000

SECTION 5
OTHER TOPICS RELATED TO BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE

METHOD FOR SPATIAL REGISTRATION OF THE EXPRESSION PATTERNS OF DROSOPHILA SEGMENTATION GENES USING WAVELETS

Kozlov K. N., Myasnikova E. M., Samsonova M. G., Reinitz J., Kosman D.

Keywords: gene expression, wavelet basis function, Drosophila

Resume

In this poster we describe how a 1D wavelet-based registration technique is applied to Drosophila segmentation gene expression patterns. Each segmentation gene's expression pattern represents a collection of "domains", each of which is a region of expression containing one concentration maximum. The images are registered by Fast Redundant Dyadic Wavelet Transform with continuous basis, used as a method for extracting ground control points. We have chosen wavelet basis functions which extract information about the first derivative of the input signal. This allows to use the input signal extrema as ground control points. The choice of the transform is motivated by such its basic properties, as noise reduction and good localisation of characteristic features. The method was applied to the set of 429 Drosophila embryos attributed by visual inspection to 8 temporal classes according to their stage of development. Embryos taken from one temporal class, i.e. of approximately the same age, have been subjected to the space registration against the common expression domain. The registration results show very good accuracy especially for embryos at late stages of development. A full description of this method has just been published in Computational Technologies 5:112.

MOLECULAR-GENETICAL MECHANISMS OF THYROID SYSTEM REGULATION: DESCRIPTION IN TRRD AND GeneNet DATABASES

Suslov V.V., *Ignatieva E.V., Osadchuk A.V.

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: eignat@bionet.nsc.ru
*Corresponding author

Keywords: databases, ES-TRRD, thyroid system, transcription regulation

Resume

Motivation:

The progress in experimental methods that is evident to worldwide scientific community initiates rapid accumulation of experimental data, which, on the one hand, presents a broad field for theoretical analysis, but, on the other hand, hinders this process extremely. This paradox appears due to enormous volume of the stored information and to enhancement of narrow specialization in science. One of possible solutions of above problem is development of digital databases enabling the specialists of various profiles to get an access to huge bulk of information.

Results:

In the ES-TRRD database, an information on thyroid system genes regulation is represented in a formalized form. In the GeneNet database, in a formalized form is given the general scheme of endocrine regulation of thyroid system.

Availability:

ES-TRRD database is available through the Internet via the address http://www.bionet.nsc.ru/trrd/es-trrd/.

Introduction

At present, rapid accumulation of experimental data on molecular-genetical mechanisms of physiological function regulation in vertebrates takes place. However, up to now, neither effort was made for accumulation in a digital form of the already available knowledge with the goal of these data further systematization and the following computer-assisted analysis. The database TRRD (Transcription Regulatory Regions Database), developed at IC&G SB RAS, enables to store experimental data on gene transcription regulation in a formalized form [Kolchanov N.A. et al., 1999], which significantly lightens their subsequent analysis. The other database, GeneNet, also developed in IC&G SB RAS, helps to represent the interrelations between genes, protein products, and hormones, which participate in realization of physiological functions, in a graphical representation viewed as gene networks [Kolpakov F.A. et al., 1998]. In the present paper, an informational contents of TRRD and GeneNet sections, devoted to thyroid system regulation, are described.

Methods

Accumulation of information in the section ES-TRRD (Endocrine System - Transcription Regulatory Regions Database) was produced on the base of annotating of scientific publications by applying the format adopted in the TRRD database. The input of information into TRRD database was made by using the original software program for data input [Kolchanov N.A., et al., 1999], input into GeneNet database – by applying original data input program [Kolpakov F.A. and Ananko E.A. 1999].

Results

In TRRD database, in the subsection ES-TRRD, an information is collected about regulation of 15 genes of thyroid system. This group consists of the genes encoding hormones or their precursors, receptors, enzymes, and transcription factors (see Table 1). In this subsection, 18 regulatory regions (promoters, enhancers, etc.) are described, along with about 60 transcription factor binding sites. This information is collected on the basis of annotating of 38 scientific papers.

According to the data stored in ES-TRRD, about 18 various transcription factors control gene expression of thyroid system. Among them are the well-known factors such as CREB, ATF-2, Pit-1, Pax-8, TR (generalized denotation of thyroid hormone receptors), together with the less studied protein components of a cell, e.g., TIF (TSHR-insulin responsive factor), NTF-1 (NIS TSH-responsive factor-1) (see Table 1).

Regulatory regions of many genes are studied enough detailly, however, to a different extent. For example, due to published data, regulatory region of thyroglobulin gene in rat (A00088) is limited by promoter region, it contains 177 bp and includes transcription factors binding sites TTF-2 and TTF-1 (Fig.1a). The similar structure has the regulatory region (not shown in Fig. 1) of rat thyroid peroxidase gene (A00979) [Sato K. and Di Lauro R., 1996]. From the other hand, regulatory region of human thyrotropin beta-subunit gene (Fig.1d) occupies 1243 bp and, except promoter, includes silencers (-1200/-48, -420/120) and regulatory region within exon 1. These regulatory regions contain binding sites of such transcription factors as Oct-1, Pit-1, TR.

Notably, binding sites of the factor TR (Fig.1d, +1/+6, +13/+18, +31/+36) form two composite elements. The composite element (CE), by definition, is represented by two sites acting as a single functional unit. Both composite elements located in T3-inhibitory region (Fig.1d, inhibitory region, +1/+43), are related to each other. Mutation in any of three sites binding TR factor prevents totally to inhibitory effect of T3 to the gene transcription [Bodenner D.L., et al., 1991].

Regulatory region of rat thyrotropin receptor gene (A00962) is located between positions -880 and –80 relatively translation start. In this regulatory region, the most of well-characterized sites functions within the content of four composite elements (namely, the sites –880/-865 and -189/-174, -220/-193 and -189/-174, -220/-193 and –165/-132, -165/-132 and –116/-80). It is interesting to note that one and the same site may enter the content of several composite elements. Thus, promoter region of the rat thyrotropin receptor gene (Fig.1c, -880/-80), is organized as a cluster of at least four intersecting composite elements. Besides, in regulation of TSHR, SSB proteins take part [Shimura H., et al., 1993], whereas the GABP factor is sensitive to DNA methylation (Fig.1c, Fig.2) [Yokomori N., et al., 1998]. Such a complex structure is likely necessary for co-ordination of insulin/IGF-dependent regulation with the TSH/cAMP-dependable regulation (Fig. 2) and, hence, for establishment of tissue-spesificity.

Murine preprothyrotropin-releasing hormone gene (ppTRH) (Fig.1b) is interesting by dispersed nature of constituent functionally related elements. It was demonstrated that inhibition of gene expression by thyroid hormones is realized under the participance of two regulatory elements, the binding site of TR factor (-62/-47) and regulatory element (+12/+47). According to experimental data on gel shift retardation, the regulatory element (+12/+47) does not bind to TR (Fig.1b). However, experiments on transfection by plasmid constructions of different length give evidence that both regulatory elements function as a united integrity, because gene expression inhibition by thyroid hormones is realized only in case the plasmid with the DNA fragment carries both elements. [Satoh T., et al., 1996].

A section of GeneNet database denoted as "Thyroid system" and devoted to molecular bases of endocrine regulation of thyroid system. The section describes the main fragment of a gene network dealing with hypothalamic, hypophysial, and thyroidal interactions, which control the central regulation of thyroid hormone synthesis. The section includes the description of 9 genes, 27 proteins, and 52 interrelations between theobjects (see Fig. 2). The data accumulated in this section were obtained on the basis of annotating of 30 research articles.

The negative regulation contour consisting of hypothalamic, hypophysial, and thyroidal interrelations is realized through

Table 1. Genes of thyroid system: content of the ES-TRRD database.
In the square brackets, the poorly studied factors are given.

Name of a gene	TRRD accession number	Orga-nism	Transcription factors acting in gene expression regulationr
Genes, encoding hormones			
thyroglobulin (TG)	A00088	rat	Sp-1, TR/RXR, TTF-1, TTF-2, USF/MLTF
glycoprotein hormone alpha-subunit (thyrotropin alpha-subunit)	A00056	human	GSEB1, GATA, CREB, GR, TR
thyrotropin beta-subunit	A00999	rat	TR, Pit-1/GHF-1
thyrotropin beta-subunit	A01002	mouse	TR/RXR
thyrotropin beta-subunit	A01001	human	TR, TRH, Oct-1, Pit-1/GHF-1
preprothyrotropin-releasing hormone	A00851	mouse	TR, EGF
preprothyrotropin-releasing hormone	A00998	rat	EGF
Genes, encoding receptors			
thyroid hormone receptor B2	A01003	mouse	Pit-1/GHF-1
thyrotropin receptor	A00962	rat	CREB, ATF-2, CREB-BR, ICERII, ICERIIgamma, GABPalpha/GABPbeta1, [SSB protein], [TIF (TSHR-insulin responsive factor)]
Genes, encoding transcription factors			
thyroid transcription factor-1	A00995	rat	TTF-1
thyroid transcription factor-1	A00996	black rat	HOXB3, HOXD3
Genes, encoding enzymes			
Na+/I- symporter	A00987	rat	TTF-1, Pax-8, [NTF-1 (NIS TSH-responsive factor-1)]
Na+/I- symporter	A00994	human	TTF-1, Pax-8
thyroid peroxidase	A00979	rat	TTF-1, HNF-3B, TTF-2
thyroid peroxidase	A00997	human	TTF-1

the action of 3 signal molecules: thyrotropin-releasing hormone (TRH), thyrotropine (TSH), and triiodothyronine (T3) (one more thyroid hormone, thyroxin (T4), is not shown for simplicity).

Except these molecules, at the diagram (Fig. 2), are shown insulin, IGF, and EGF. These molecules do ot refer to the central regulation of thyroid system, but they are necessary for it, because they provide peripheral regulation. In particular, the coordinated functioning of thyroid system with the other endocrine regulation systems is provided with their help.

TRH, TSH, T3, insulin, IGF, and EGF represent the extra-cellular regulation level at the diagram (Fig. 2). Hormone receptors (TRHR, TSHR, TR, etc.) serve as connection link between the upper-cellular and inner-cellular regulation levels. Inner-cellular regulation level includes genes and transcription factors (see Table 1). Besides, some receptors act as transcription factors by themselves (see Table 1).

At the diagram, the inner-cellular level of thyroid hormone regulation in thyroid is given in more details. It consists of the subsystem of the genes controlling biosynthesis of thyroid hormones and the gene responsible for accepting external signals, the thyrotropin receptor gene (TSHR, A00962).

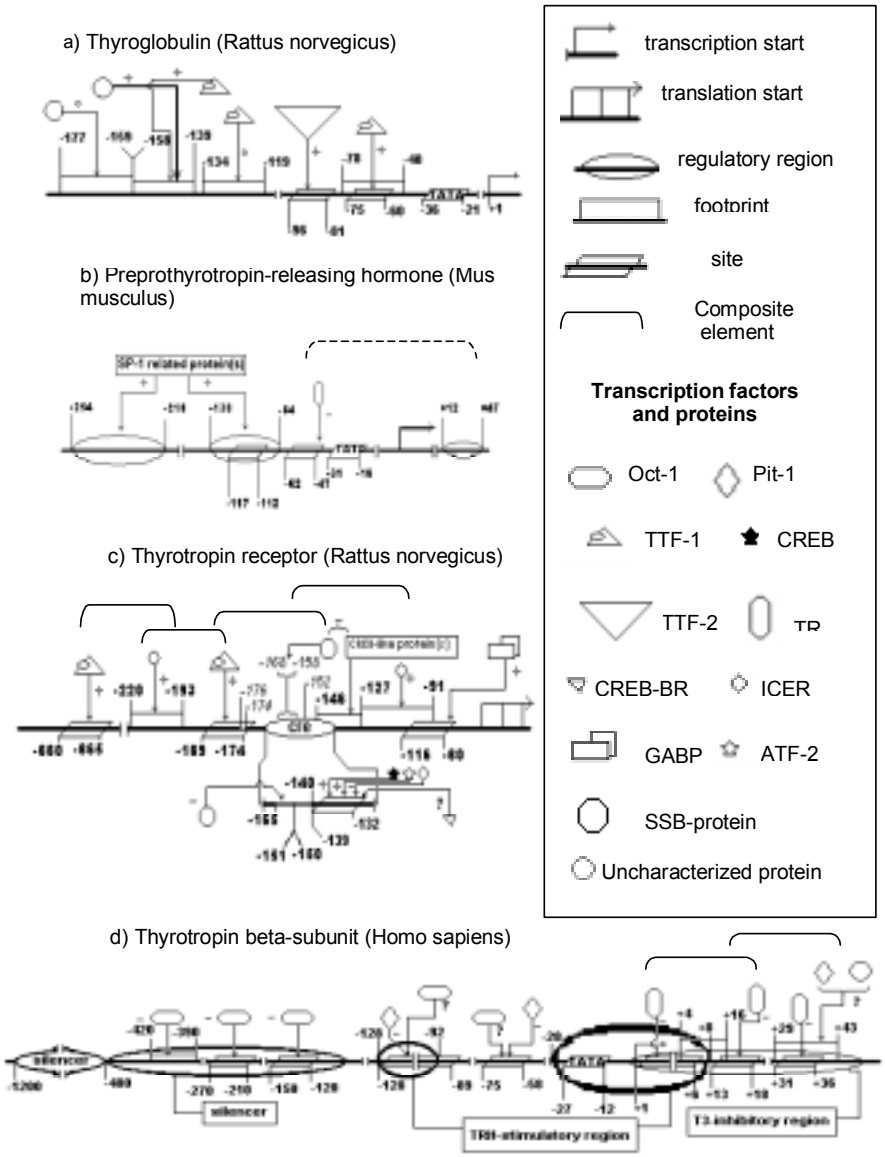


Figure 1. The schemes of regulatory regions of some genes, which are referred to thyroid system and contained in the ES-TRRD database. A small oval in denotation of a protein means that this protein is phosphorylated. By light numbers are denoted positions sensible to DNAse, which splits single-stranded DNA. Bold arrows directed from transcription factors denote that the other factors bind to the site only in case this factor is absent. In Fig. 1b, by the dash line of a composite element, the functionally related elements are shown.

Sub-system of genes, determining biosynthesis as it is, consists of TATA-containing genes: NIS (A00987), TPO (A00979), TG (A00088). Coordinated regulation of these genes is assured by combination of TTF-1, TTF-2, and Pax-8. TSH–stimulation of this system is governed through NIS gene regulation (A00987) by means of poorly studied protein component NTF-1.

TSHR (A00962) expression depends upon duration of TSH action. Of special role in receptor expression regulation are the proteins CREB and ICER, which interact with one and the same site – cAMP-dependent element, (CRE, -165/-132), given in Fig. 1c. Short-run (1-2 h) action of TSH on follicular thyroid cell causes binding of CREB to the site CRE, which, in turn, stimulates TSHR synthesis. When the action of TSH is more durable (more than 4 h), the other protein, ICER, interacts with this site, this process inhibiting TSHR synthesis [Saji M., et al., 1992], [Lalli E. and Sassone-Corsi P., 1995]. Except this classical pathway, the inhibition may be exerted by the CREB-family proteins, with involvement of the SSB proteins [Shimura H., et al., 1993]. How this way of inhibition agrees with ICER inhibition is still unclear.

Discussion

We have accumulated an information about molecular mechanisms of thyroid system regulation. In future, we plan to make further accumulation of information in TRRD and GeneNet databases according the following directions.

As for the central regulation of thyroid system, it is necessary to make a more sophisticated description of regulation at the inner-cellular level in hypothalamus (1) and hypophysis (2); to introduce an information about the action of thyroid hormones directly on thyroid (3). Moreover, it is possible to include the data on regulation of

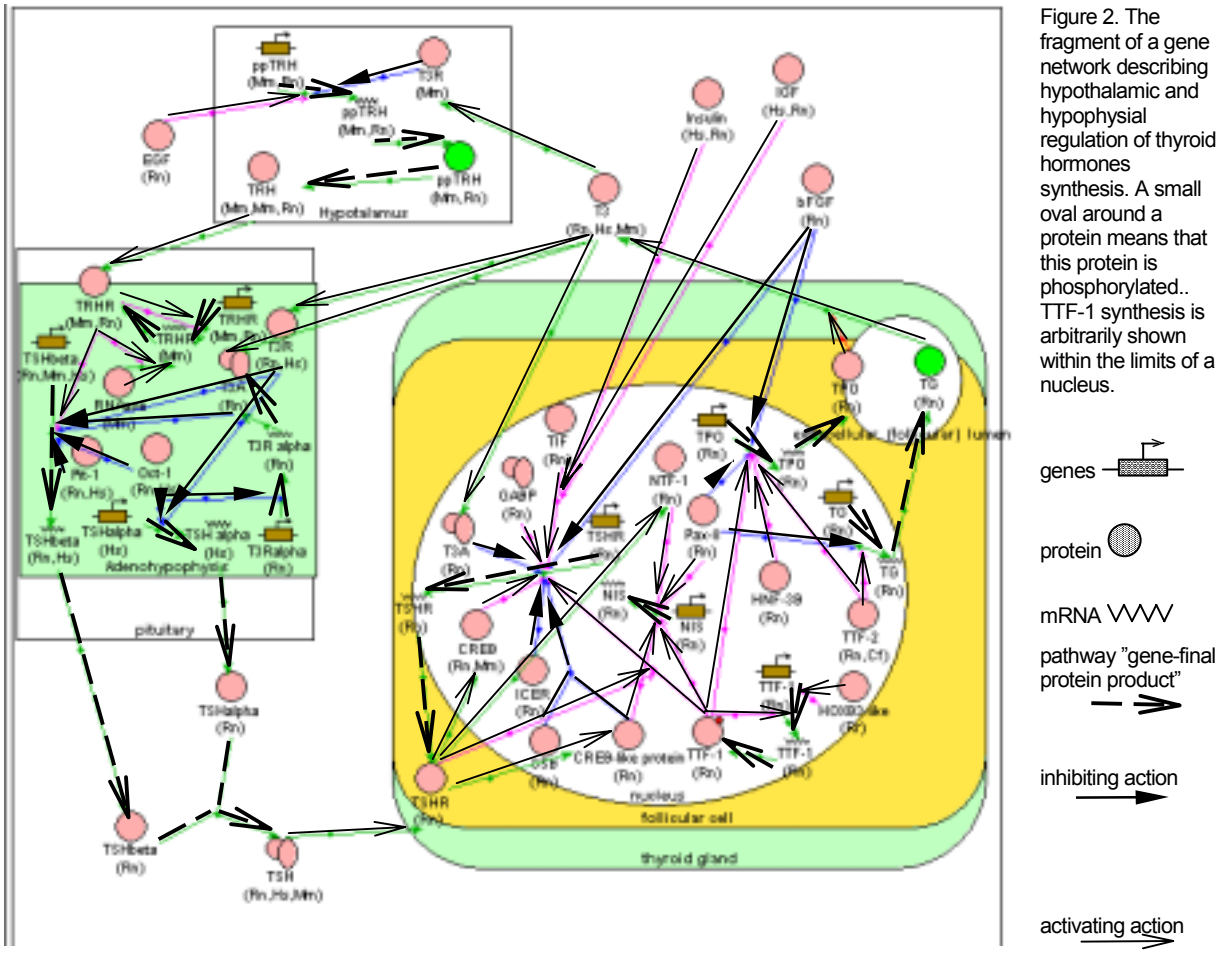


Figure 2. The fragment of a gene network describing hypothalamic and hypophysial regulation of thyroid hormones synthesis. A small oval around a protein means that this protein is phosphorylated.. TTF-1 synthesis is arbitrarily shown within the limits of a nucleus.

thyroid system at the upper-hypothalamus levels (4).

As for peripheral thyroid system regulation, it is necessary to characterize the interrelations of thyroid system and endocrine regulation system, which are functionally related (5). These interrelations should be supported by the presence of common transcription factors (EGF, IGF) and hormones, or their subunits (insulin, TSHalpha). A particular area of research is produced by accumulation of information on thyroid hormones binding sites located within regulatory sites of genes referring to the other endocrine systems, in particular, the system of sex hormones synthesis (6). In addition, it will be worthwhile to collect an information in TRRD database about binding sites of such wide-spread factors as CREB or TTF-1 (7).

Acknowledgements

This work is supported by Russian Foundation for Basic Research (grants Nos. 98-04-49479, 00-04-49229, 00-04-49255, 00-07-90337, 98-04-49452). The authors are grateful to I.V. Lokhova for the help with achieving literature sources and to G.V. Orlova for translation of the manuscript into English.

References

1. D.L. Bodenner, M.A. Mrokzynski, B.D. Weintraub, S. Radoick, F.E. Wondisford, "Detailed functional and structural analysis of a major thyroid hormon inhibitory element in the human thyrotropin beta-subunit gene" J.Biol.Chem. **266,** 32, 21666-21673(1991).
2. N.A. Kolchanov, E.A. Ananko, O.A. Podkolodnaya, E.V. Ignatieva, I.L. Stepanenko, O.V. Kel-Margoulis, A.E. Kel, T.I T.I T.I. Merkulova, T.N. Goryachkovskaya, T.N. Busigina, F.A. Kolpakov, N.L. Podkolodny, A.N. Naumochkin, A.G. Romashchenko, "Transcription Regulatory Regions Database (TRRD): its status in 1999" Nucleic Acids Res. **27,** 303-306(1999).
3. F.A. Kolpakov, E.A. Ananko, "Interactive data input into the GeneNet database" Bioinformatics **15,** 713-714(1999).
4. F.A. Kolpakov, E.A. Ananko, G.B. Kolesov, N.A. Kolchanov, "GeneNet: a gene network database and its automated visualization" Bioinformatics **14,** 529 – 537(1998).
5. E. Lalli, P. Sassone-Corsi, "Thyroid-stimulating hormone (TSH)-directed induction of the CREM gene in the thyroid gland participates in the long-term desensitization of the TSH receptor" Proc.Natl.Acad.Sci.USA. **92,** 21, 9633-9637(1995).
6. M. Saji, T. Akamizu, M. Sanchez, S. Obici, E. Avvedimento, M. E. Gottesman, D. E. Kohn, "Regulation of thyrotropin receptor gene expression in the rat FRTL-5 thyroid cells" Endocrinology **130,** 1, 520-533(1992).
7. K. Sato, R. Di Lauro, "Hepatocyte nuclear factor 3beta participates in the transcriptional regulation of the thyroperoxidase promoter" Biochem.Biophys.Res.Comm.,**220,** 1, 86-93(1996).
8. T. Satoh, M. Yamada, T. Iwasaki, M. Mori, "Negative regulation of the gene for the preprothyrotropin-releasing hormone from the mouse by thyroid hormone requires additional factors in conjunction with thyroid hormone receptors" J.Biol.Chem. **271,** 44, 27919-27926(1996).
9. H. Shimura, S. Ikuyama, Y. Shimura, L.D. Kohn, "The cAMP response element in the rat thyrotropin receptor promoter. Regulation by each decanucleotide of a flanking tandem repeat uses different, additive, and novel mechanisms" J.Biol.Chem. **268,** 32, 24125-24137(1993).
10. H.J. Steinfelder, P. Hauser, Y. Nakayama, S. Radovick, J.H. McClaskey, T. Taylor, B.D. Weintraub, F.E. Wondisford, "Thyrotropin-releasing hormone regulation of human TSHB expression: role of a pituitary-specific transcription factor (Pit-1/GHF-1) and potential interaction with a thyroid hormone-inhibitory element" Proc.Natl.Acad.Sci.U S A **88,** 8, 3130-3134(1991).
11. N. Yokomori, M. Tawata, T. Saito, H. Shimura, T. Onaya, "Regulation of the rat thyrotropin receptor gene by the methylation-sensitive transcription factor GA-binding protein" Mol.Endocrinol. **12,** 8, 1241-1249(1998).

COMPARISON OF MILK PROTEIN KINASE WITH SECRETORY IMMUNOGLOBULIN A POSSESSING PROTEIN KINASE ACTIVITY

Akimzhanov A. M., Semenov D. V., *Nevinsky G.A.

Institute of Bioorganic Chemistry SB RAS, Novosibirsk, Russia
e-mail: Nevinsky@niboch.nsc.ru
*Corresponding author

Keywords: human milk protein kinase, catalytic antibodies with protein kinase activity

Resume

Motivation:

The ability of sIgA purified by several affinity chromatographies from milk of healthy mothers to phosphorylate selectively serine residues of casein in the presence of [γ-^{32}P]ATP was shown to be an intrinsic property of the antibodies [1-3]. It is a first example of the catalytic activity of natural antibodies with synthetic activity. In addition, the findings speak in favor of a possibility of Abzymes generation in tissue of healthy mammals. Therefore, it was very interesting to compare sIgA-protein kinase with any possible enzymes of human milk possessing protein kinase activity.

Results:

Here we present evidence that in addition to sIgA-protein kinase human milk contains only one protein kinase of non-immunoglobulin nature (PK) of an usually high molecular mass (~75 kDa). This enzyme phosphorylates selectively serine residues of only casein and in contrast to catalytic antibodies can not catlyze phosphorylation of other proteins of human milk. In addition, sIgA utilizes as a substrate not only ATP but also other different NTP and dNTP with comparable efficiency, while PK can use only ATP. Thus the biochemical properties of sIgA-kinase and PK are quite different. The detection of milk proteins possessing phosphorylating activity has a special interest: it is known that phosphorylation of proteins plays a fundamental role in the regulation of key physiological processes.

Introduction

Casein phosphorylation is the key stage during posttranslational protein modification occurring in the Golgi apparatus of mammary gland secretory cell. At the present moment, protein kinases responsible for casein phosphorylation *in vivo* have been found in bovine mammary gland, rat mammary gland, and in the Golgi enriched fractions of lactating mammary gland of guinea-pig. These enzymes are characterized by specific phosphorylation of only casein. The rate of phosphorylation of casein by these enzymes increases in the presence of Ca^{2+}, Mn^{2+}, and Mg^{2+} ions.

In the literature there is no available data concerning any protein kinases of human milk except our data on protein kinase activity of small fractions of sIgA antibodies [1-3]. In order to compare and to distinguish protein kinase activities of immunoglobulin and non-immunoglobulin nature in present article we have analyzed for the first time simple protein kinases of human milk.

Methods

Protein kinase Purification. After removing the lipid phase, milk was loaded on a heparin Sepharose column equilibrated by buffer A: 50 mM Tris-HCl, pH 6.8, 0.1 mM PMSF, 1% Triton X-100 and 0,1 mM 2-mercaproethanol. Then proteins adsorbed were eluted with a gradient of concentration of NaCl (0 – 1 M) in buffer A. Column fractions containing protein kinase activity were combined and dialyzed against 10 mM Tris-HCl, pH 6.8.

The fractions from heparin-Sepharose were chromatographed on DEAE-cellulose column equilibrated by buffer A. The protein was eluted with a gradient of concentration of NaCl (0 – 0.7 M) in buffer A. The active fractions were combined and gel-filtered on Toyopearl HW-55 using buffer A, containing 0.5 M NaCl. .

Activity assay. All reaction mixtures contained optimal concentrations of the components: 10 mM Tris-HCl buffer (pH 6.8), 30 mM $MgCl_2$, 0.1 mM EDTA, 10 mg/ml human milk casein, and [γ-^{32}P]ATP (0.5 mBq). Phosphorylation of milk proteins was performed at 37^{o} C for 5 - 30 min. The PK-activity was analyzed by SDS/PAGE and autoradiography according to [1-3].

Results and discussion

First, we have purified PK by affinity chromatography of milk proteins on heparin-Sepharose (Fig. 1). The protein kinase was eluted as one symmetrical peak, which can indicate in favor of existence of only one human milk protein kinase of non-immunoglobulin nature. According to data of SDS-PAGE and immunological analysis this preparation contained lactoferrin (LF; ~80 kDa) as a major extrinsic protein.

PK was then separated from other proteins by chromatography on DEAE cellulose (Fig. 2). Protein kinase activity was eluted as three peaks, which may be connected with the enzyme interaction with LF existing as several oligomeric forms containing from 1 to 4 molecules of the protein. Even after this chromatography preparation of PK contained remarkable amount of LF. It should be mentioned that such preparations of PK were very stable and they retained about 80-90 % of the initial activity during several months of keeping.

Additional purification of PK was achieved by gel-filtration of the protein (Fig 3). As one can see from Fig. 3, protein kinase activity corresponds to a protein with molecular mass about 70-75 kDa. Since the major form of LF is a tetramer, this protein was effectively separated from PK by gel-filtration. At the same time, in contrast to PK purified by DEAE cellulose, after removing of LF the enzyme become to be very unstable.

Thus, the activity of PK was decreased up to 30 % after 3 days and PK became to be essentially inactive after keeping of its solution at – 20°C during a week. This phenomenon may be a result of a significant stabilization of PK due to its interaction with LF in micelle-like complexes and destruction of such complexes after removing of LF.

According to electrophoretic data after all steps of purification the protein obtained contains two polypeptides with molecular masses about 72 and 75 kDa (Fig. 4).

One can not exclude that one of these protein bands corresponds to additional protein, which can not be separated using the above described scheme of PK purification. However, it is also possible that 72 kDa polypeptide is a product of partial degradation of 75 kDa PK during the enzyme purification. In any case, gel-filtration and electrophoretic data clearly demonstrate that in contrast to many known oligomeric mammalian PKs, human milk PK consists of one polypeptide chain and has high molecular mass similar to that for PK of Golgi apparatus of guinea-pig (70 kDa).

Human milk PK was shown to phosphorylate Ser residues of only casein. In contrast to sIgA-PK, this enzyme can not phosphorylate other milk proteins. In addition, optimal concentration of $MgCl_2$ (30-50 mM) for human milk PK is about 10-500 times higher than that for sIgA-PK (0.1 mM) and for mammalian PKs of the known first and second types (1 - 5 мМ). The K_m value for ATP in the phosphorylation reaction was estimated to be about 60 μM, while this value for ATP in the case of sIgA-PK is lower (0.1-1.0 μM) [2].

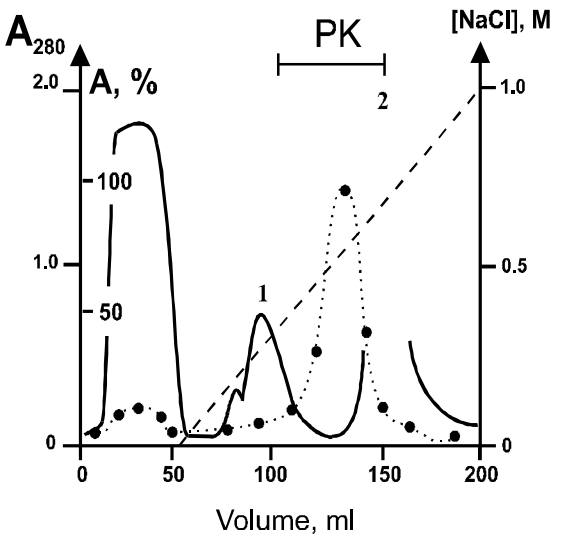


Figure 1. Affinity chromatography of human milk proteins on heparin-cepharose.: (--) - absorption at 280 nm, (•) - relative protein kinase activity.

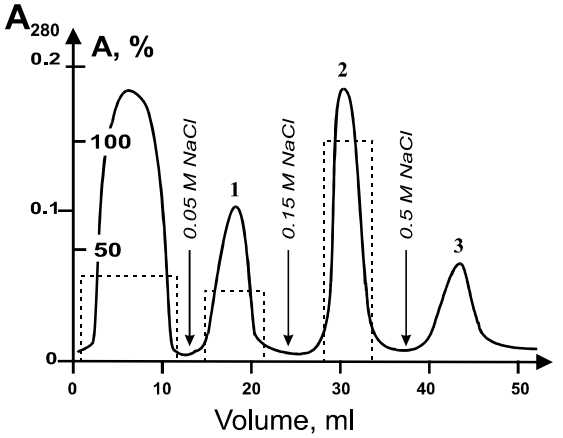


Figure 2. DEAE-cellulose chromatography of human milk protein kinase previously purified on heparin-Sepharose.: (—) - absorption at 280 nm, (- -) – relative protein kinase activity.

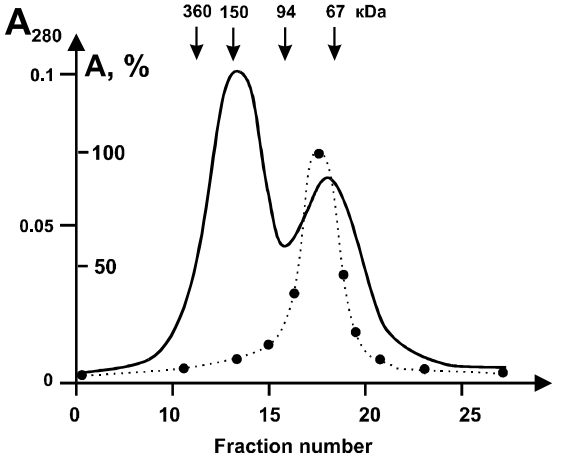


Figure 3. Purification of human milk protein kinase by gel-filtration on Toyopearl HW-55.:(—) - absorption at 280 nm, (•) - relative protein kinase activity.

In contrast to human milk PK and other different mammalian PKs, sIgA incorporates [^{32}P]phosphate into serine residues of 10-15 different milk proteins and utilizes as a substrate not only ATP but also other NTP and dNTP with comparable efficiency: ATP (100 %), dATP (70-80 %), GTP and dGTP (200-300%), UTP and dTTP (30-60%) [2].

In addition, the proteolytic hydrolysis of the casein phosphorylated by milk kinases of non-immunoglobulin nature resulted in formation of many very short labeled polypeptides, when digestion of the casein phosphorylated by sIgA fractions led to preferential formation of two high molecular masses labeled polypeptides (data not shown). It means that phosphorylation of casein by sIgA and by normal milk PK occurs in different sites of the casein.

In literature there was no available data concerning phosphorylation of proteins in human milk as well as about catalytic activity of sIgA (IgA) or natural Abzymes with synthetic activity. Our data demonstrate that in mothers milk in the presence of ATP an Abzyme-dependent phosphorylation of various proteins and simple PK-dependent phosphorylation of casein takes place.

The above results show that a possible way of Abs and Abzymes generation is still not completely understood and classical immunological concept about a role and function of Abzymes in the mammals probably will be supplemented in the future.

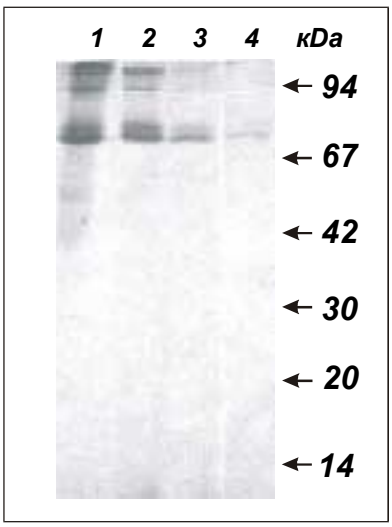


Figure 4. SDS-polyacrylamide gel electrophoretic analysis of fractions obtained by gel-filtration onToyopearl HW-55 (Fig 3): 1 – initial preparation of PK before gel-filtration, 2 – fractions 10-15, fractions 15-18, fractions 19-25. Gel was stained with Coomassie.

Acknowledgments

This work was supported by grants from the Russian Foundation for Basic Research (No. 98-04-49719), Russian Ministry of General and Professional Education (" Studies in Chemical Technologies", No. 98-8-3.2-69) and grants for young scientists from the Siberian Branch of Russian Academy of Sciences.

References

1. Nevinsky G.A., Kit Yu. Ya., Kit Yu., Semenov D., Nevinsky G. (1995) Protein Kinase Activity of sIgA antibodies from human milk: CATALYTICALLY ACTIVE ANTIBODIES IN NORMAL HUMANS?, *Molecular Biology*, **29**, 519-526.
2. Nevinsky G.A., Kit Yu. Ya., Semenov D.V.,. Buneva V.N.(1998) Secretory Immunoglobulin A from Human Milk Catalyzes Reaction of the Milk Proteins Phosphorylation*. Appl. Biochem. Biotechnol.,* **75**, 77-91.
3. Kit Yu.Ya., Semenov D.V. and Nevinsky G.A.(1996) Phosphorylation of different human milk proteins by immunoglobulin A. *Bioch. Mol. Biol. Internat.,* **39**, 521-527.

UNEXPECTED HIGH AFFINITY OF SOME VERY SHORT NONSPECIFIC OLIGONUCLEOTIDES FOR HUMAN DNA TOPOISOMERASE I

Bugreev D.V, *Nevinsky G.A.

Institute of Bioorganic Chemistry SB RAS, Novosibirsk, Russia
e-mail: nevinsky@bioch.nsc.ru
*Corresponding author

Keywords: DNA topoisomerase I, short nonspecific oligonucleotides, inhibition

Resume

Motivation & Results:

Eukaryotic DNA topo I is an abundant nuclear enzyme that control and modify the topological states of negatively and positively supercoiled DNA by introducing transient single-stranded nicks in the nucleic acid backbone. DNA topoisomerase I have been found to affect a number of vital biological functions including the DNA replication, transcription, recombination, chromosome dynamics, have been implicated in others and is the pharmacological targets of a number of antitumor drugs. Thus the butter understanding of the enzyme function molecular mechanism is important from fundamental and applied points of view.

The enzyme affinity for specific sequence oligonucleotides (ODNs) corresponding to scissile strand (5-15 nucleotide units in length) is about 2-4 orders of magnitude higher then that for nonspecific ODNs of the same length. The affinity of specific very short ODNs (2-4 nucleotides) are also higher than those of nonspecific ones. However, such very short nonspecific ODNs have unpredicted high affinity for the enzyme. A possible reason of this phenomenon is discussed.

Introduction

Topo I is a sequence-specific enzyme. This enzyme protects both DNA strands over a 20-base binding region in which the cleavage site is centrally located ([1-3] and refs therein). Enzyme-mediated DNA cleavage requires an asymmetric region encompassing the cleavage site; it has been established that the minimal region consists of nine nucleotides on the scissile strand and five nucleotides on the non-scissile strand, with the majority of the nucleotides situated upstream of the cleavage site (Fig.1).

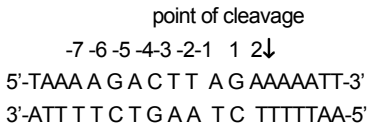


Figure 1. Position of point of cleavage

Since cleavage in DNA occurs only at sites situated in regions with potential for intrastrand base-pairing due to distal complementary sequences [4-6, and references therein], Topo I has been considered to be a ds-specific enzyme which can only recognized DNA molecules of reasonable length, and it was assumed that topo I cannot interact with short single-stranded ODNs.

At the same time we have shown recently that both 5-15 mers specific and non-specific ODNs have a high affinity for the enzyme and are efficient inhibitors of topo I-dependent relaxation of scDNA [1-3]. In this article we examined whether human topo I can bind to very short ss of specific and nonspecific sequences, and compared a relative affinity of the enzyme for such oligonucleotides.

Methods

Topo I activity was measured by assessing relaxation of sc colE1 DNA at 30° C. The reaction mixture (20 μl) contained: 50 mM Tris-HCl (pH 8.0), 70 μg/ml BSA, 0.5 mM DTT, 0.5 mM EDTA, 15% glycerol, 130 mM NaCl, 10 μg/ml colE1 scDNA and 1 unit of topo I. ODNs were used at different concentrations. Samples were incubated for 15 min and inhibition of topo I-dependent relaxation of scDNA by ODNs were estimated by agarose gel electrophoresis as in [1-3]. Ethidium bromide-stained gels were photographed and the films were scanned. The activities of topo I and the IC_{50} values for ODNs were estimated from the scanning densitometry data.

Results and discussion

It is commonly supposed that enzymes recognizing long dsDNA are not capable of interacting effectively with minimal ligands like dNMP, or short or long ss ODNs [4-5]. We analyzed the efficiency of interaction of several replication, repair, integration, and restriction enzymes with dNMPs (orthophosphate and deoxyribose) and

showed that all these enzymes recognizing any specific or nonspecific long DNA molecule are capable of binding with a definite affinity to dNMPs and to their structural components modeling the specific elements of DNA nucleotide units (for review see [4-5]). It should be mentioned that the efficiency of the interaction of enzymes with dNMPs as minimal ligands can be lower or higher, but in most cases the ratio of K_d values for dNMP and for the long DNA accounts to discriminate between of short and long DNA-ligands. It was shown that virtually all nucleotide units within the DNA binding cleft interact with the enzymes, and that high affinity is mainly (5-8 orders of magnitude) is provided by many weak additive interactions between the enzyme and various structural elements of many DNA nucleotide units.

For all the enzymes studied, the logarithmic dependencies of the K_d values for ss and ds $d(pN)_n$ ODNs upon their length (n) appear to be linear for 0-1 $\leq n \leq$ 7-20 (depending on the enzyme); the extrapolation of most dependencies to n = 0 gives the K_d value coinciding with the experimental K_d values for orthophosphate, which is one of the minimal ligands of all the enzymes analyzed. These data provide evidence for the additivity of the Gibb's free energies characterizing interaction of the individual $d(pN)_n$ units with these enzymes.

The values of *f* factors characterizing the increase in enzyme affinity for various $d(pN)_n$ for one-unit-increase in their length were evaluated from the slopes of the linear parts of the logarithmic curves. The absolute values of *f* factors depended on the enzyme, as well as on the type of bases of homo-$d(pN)_n$; The K_d values accounting for the efficiency of interaction between the enzyme and one unit of ss DNA are equal to the reciprocals of these factors (1/*f*). It should be emphasized that the K_d values for any homo-ODNs in the case of most enzyme studied can be obtained by multiplying K_d for the minimal ligand (most commonly dNMP or P_i) by $K_d = 1/f$ for each of the mononucleotide units, according to a geometrical progression: $K_d[(pN)_n] = K_d[(P_i)] \times [1/f]^n = K_d[(dNMP)] \times [1/f]^{n-1}$ ($1 < n \leq$ 7-20, depending on the enzyme).

The increase in the affinity of the enzymes for ss and ds DNA due to their interaction directly with one of the internucleoside phosphate groups (electrostatic factor *e*) was evaluated from the slopes of the logarithmic dependencies of lg *f* versus the relative hydrophobicity of the bases; $f = e \bullet h$, where *h* is the hydrophobic factor reflecting the relative hydrophobicity of the C, T, G and A bases of different $d(pN)_n$.

It was shown that the interaction of any investigated sequence-dependent and sequence-independent enzymes with nonspecific ss and ds DNA can be described using the common algorithm: $K_d[d(pN)_n] = K_d[(P_i)] \times (e)^{-n} \times (h_C)^{-l} \times (h_T)^{-m} \times (h_G)^{k} \times (h_A)^{-g}$, where l, m, k and g - the number of corresponding bases in DNA ligand. When passing from one enzyme to another only the values of *e (1.3-1.7)* and h_N factors (1- 1.7) change.

Interestingly, in contrast to the all above mentioned enzymes the interaction of Topo I with very short nonspecific ODNs can not be described using the descending progression presented above. Our data on Topo I speak in favor of the idea that in some case the affinity of very short ODNs can be higher or lower than that for the same ODNs within a long DNA. For example, Fig. 2 demonstrates the absence of a typical linearity of lg-dependencies of the I_{50} values upon length (n) of the $d(pN)_n$ ODNs. dTMP and $d(pT)_2$ have unpredicted high affinity for the DNA binding site of Topo I.

Transition from $d(pT)_2$ to $d(pT)_3$ leads to a significant decrease of the ligand affinity (2-2.5 order of magnitude) and a further increase of $d(pT)_n$ length gives a practically linear dependence of the lgI_{50} upon *n*. As one can see from Fig. 2, the affinity of short specific $d(pN)_{3-9}$ (corresponding to cleavable chain of DNA substrate) is significantly higher than that for nonspecific oligothymidilates. At the same time a regularity of a change in the affinity for $d(pT)_n$ is a similar for that in the case of short specific ODNs (Fig. 2). The dependencies of - lgI_{50} on the length for some other short $d(pC)_n$ .and $d(pA)_n$ are also nonlinear and several d(pN)n (n= 2-4) have also abnormally high affinity for Topo I (data not shown). In addition, the character of the change of the enzyme affinity for these ODNs is different than that for $d(pT)_{2-4}$. Interestingly, that affinity of $d(pC)_2$ and $d(pA)_2$ is lower than that for $d(pT)_2$ by a factor of 2-25, while the I_{50} values for $d(pC)_{3-6}$ and $d(pA)_{4-6}$ is about 100-fold lower that those for corresponding oligothymidilates. It should be mentioned, that Topo I is the first example of enzymes interacting with nucleotide units of short ODNs in a non-additive manner.

In order to analyze a possible reason for atypical interaction of ODNs with Topo I, the following should be taken into account. According to X-ray structural data (Redinbo et al, 1998; Stewart et al, 1998), Topo I interacts mainly with many internucleoside phosphate groups of specific DNA substrates forming nonspecific electrostatic contacts and hydrogen bonds. The enzyme forms only one specific contact in the DNA minor groove: N-ε-Lys532 with O-2(-1)T (see a scheme).These contacts are necessary for strong binding of the specific sequence to the enzyme as a whole (Redinbo et al., 1998).

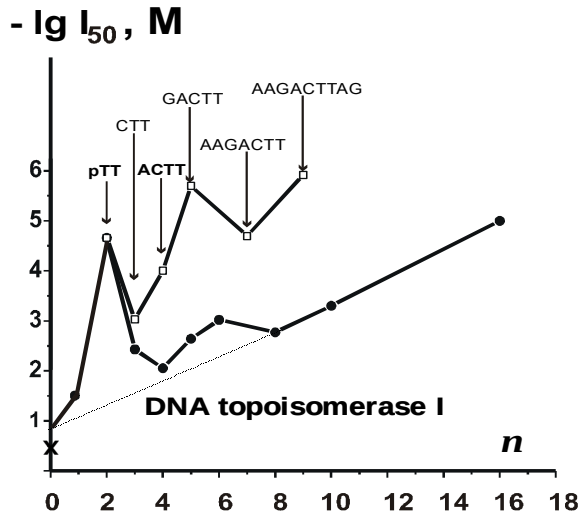


Figure 2. Dependencies of –lg I50 for single-stranded d(pT)n and specific ODNs versus their length (n) for Topo I

After primary complex formation Topo I changes the ODNs structure to the optimal conformation [1-5]. As a result preincubation increases the affinity of ODNs (factor 3.0-42.0), and for specific ODNs this factor is about 2-14 times higher than that for non-specific ODNs and a significant affinity increase occurs when complementary or partially complementary ligands are used (up to a factor of 42.0).

Therefore it is reasonable to suggest that $d(pT)_2$ can change its conformation and interact with the enzyme due to partial formation of the contacts of this dinucleotide in specific sequence AAGACTT. The affinity of $d(pC)_2$ and $d(pA)_2$ is most probably lower since these dinucleotides can not form specific hydrogen bond of T-base of specific sequence. At the same time according to X-ray analysis data (Fig. 3) the enzyme form many electrostatic contacts and hydrogen bonds with internucleoside phosphates of GACT part of the sequence (see the scheme). Consequently, one can suppose that structural features of $d(pC)_{4-6}$ and $d(pA)_{3-6}$ ODNs are more suitable for interaction of such ODNs (in comparison with $d(pT)_{3-6}$) with specific part of the enzyme recognizing GACT tetranucleotide of specific sequence.

Thus it is obvious, that nonspecific weak electrostatic interactions and hydrogen bonds between non-cognate

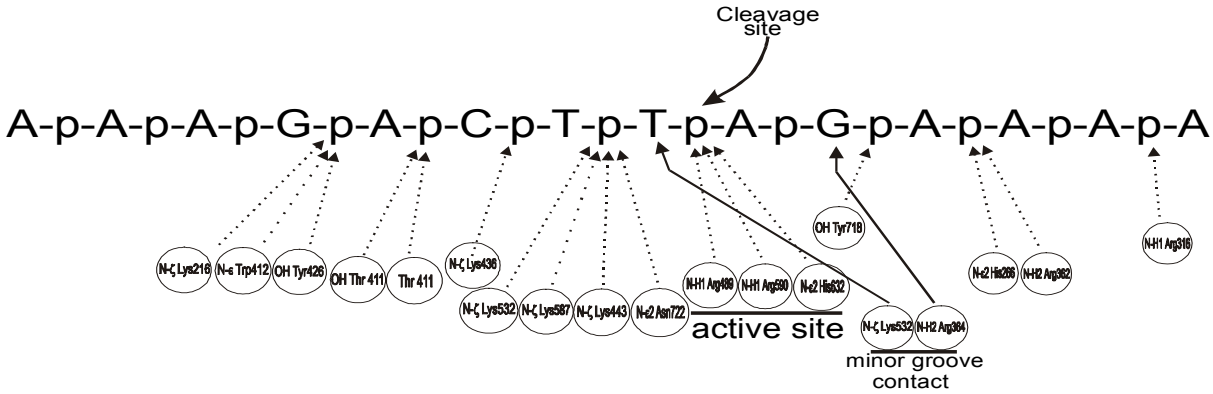


Figure 3. The scheme of contacts of aminoacids of Topo I with structural elements of cleavage chain of specific sequence (from X-ray data).

ODNs and Topo I become significantly stronger (and acquire a character of specific interactions) at the transition from nonspecific to specific ones. Therefore, it seems reasonable to suggest that the unexpected high affinity of some types of short ODNs may be a result of similar stronger interactions of the enzyme not only with specific ODNs but also with some short nonspecific ones. But within the longer ODNs these mono or dinucleotide units most probably can not interact with the enzyme in the same way and therefore in the case of $d(pT)_n$ at $n > 4$, the logarithmic dependencies become linear.

References

1. Nevinsky G.A., Bugreev D.V., Buneva V.N., Yasui Y., Nishizawa M., Andoh T.(1995) High affinty interaction of mammalian topoisomerase I with short, single and double-stranded oligonucleotides with specific sequences, *FEBS Lett.*, 368, 97-100.
2. Bugreev D.V., Vasyutina E.L., Buneva V.N., Yasui Y., Nishizawa M., Andoh T., Nevinsky G.A. (1997) High affinity interaction of mouse DNA topoisomerase I with di- and trinucleotides corresponding to specific sequences of supercoiled DNA cleaved chain. *FEBS Lett.,* 21; 18-20.

3. Bugreev D.V., Vasyutina E.L., Kolocheva T.I., Buneva V.N., Andoh T., Nevinsky G.A. (1998) Interaction of human DNA topoisomerase I with specific sequences oligodeoxynucleotides. *Biochimie.* 80, 303-308.
4. Nevinskii, G.A. (1995) An important role for weak interactions during the recognition of long DNA and RNA molecules by enzymes. *Mol. Biol. (Moscow)* 29, 16-37.
5. Bugreev DV, Nevinsky G. A. (1999) Possibilities of the method of step-by-step complication of ligand structure in studies of protein--nucleic acid interactions: mechanisms of functioning of some replication, repair, topoisomerization, and restriction enzymes. *Biochemistry (Moscow)* 64, 237-249.

EFFECT OF DNA-HYDROLYZING ANTIBODIES ON CELL DIVISION AND IMMUNORESPONSE OF EXPERIMENTAL MICE

Gal'vita A.V.[1], *Kanyshkova T.G.[1], Isaeva N. V.[2], Polosukhina D.I.[3], Nevinsky G.A[1].

[1]Institute of Bioorganic Chemistry SB RAS, Novosibirsk, Russia
[2]Institute of Cytology and Genetics SB RAS, Russia
[3]Novosibirsk State University, Russia
e-mail: kanyshk@niboch.nsc.ru
*Corresponding author

Keywords: DNA-hydrolyzing antibodies, autoimmune diseases, autoantibody cytotoxicity, antiidiotypic mechanisms

Resume

Motivation:

It is known that autoantibodies, like enzymes, can possess catalytic activity ([1-5] and refs therein). During the last several years DNA-hydrolyzing antibodies from sera of patient with several autoimmune and viral pathologies, and lymphoproliferative diseases have been isolated and extensively studied. However a mechanism of their formation and mode of DNA-abzyme action *in vivo* is rather pool understand.

Results:

Here we present evidence that DNA-hydrolyzing antibodies (Abs) from blood of patients with systemic lupus erythematosus (SLE) and multiple sclerosis (MS) possess different effect on cell division and demonstrate various autoimmune response in experimental mice.

Introduction

DNA- and RNA-hydrolyzing Abs were detected in the sera of patients with several autoimmune and viral diseases: SLE, rheumatoid arthritis, sclerodermia, Hashimoto's thyroiditis, polyarthritis, MS, viral hepatitis, and AIDS ([1-3], and refs therein). We have shown also that the milk of normal human mothers contains sIgA and IgG Abs hydrolyzing DNA, RNA [4]. At the same time, we didn't detect DNAase or RNAase activities of Abs from the sera of 50 normal humans (men and women) and in sera of patients with influenza, pneumonia, tuberculosis, tonsilitis, duodenal ulcer, and some types of cancer. The nuclease activities were shown to be an intrinsic property of auto-Abs in all the above cases [1-4].

All the data obtained demonstrate an extremely pronounced heterogeneity of polyclonal nuclease Abs from sera of autoimmune patients concerning all their kinetic characteristics, substrate specificities, and relative specific activities [2,3,5 and refs therein]. Therefore, it is very interesting to clarify a possible correlation of the above characteristics of abzymes from patients with different pathologies with their effect of deterioration or improvement of the clinical state of the patients.

Here we present for the first time the data on effect of DNA-hydrolyzing Abs from blood of patients with MS and SLE on the cell division and immunoresponse of experimental mice.

Methods and algorithms

Electrophoretically and immunologically homogeneous IgG and IgM from sera patients and human milk were obtained as described in [1]. DNAase activity of Abs was analyzed using plasmid DNA and 5'-[^{32}P]ON under optimal conditions (20 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1 mM EDTA, 37^{o} C). BALB/c mice were immunized intradermally in the footpads. The induction of DNA-hydrolyzing abzymes in mice blood were analyzed as in the case of human sera. Cytotoxicity of abzymes was studied as in [5].

Results and discussion

As yet, the important question of the biological role of catalytic Abs in the blood of patients with autoimmune diseases remains unanswered. Is it the result of function or of dysfunction of the immune system?

As one can see from the literature data, in autoimmune diseases there may be abzymes which are specific for each type of disease, such as IgGs hydrolyzing vasoactive neurotransmitter peptide in bronchial asthma [*6*] and thyroglobulin in Hashimoto's thyroiditis [Paul, S. et al. Science 1998]. According to Paul's data, anti-peptide Abs are cytotoxic and immunization of mice with anti-hormone catalytic IgGs from human sera leads to induction of experimental asthma. At first glance, one can suggest that DNA- and RNA-hydrolyzing abzymes may be considered as "non-specific abzymes", since according to our data they may be detected in the sera of patients suffering from many different autoimmune and viral diseases. Therefore such abzymes may be side- and/or

accompanied- product of the autoimmune processes. A crucial question is whether such Abs are cytotoxic and can have any important role in the etiology and/or pathogenesis of various diseases. In this connection, Sinohara at al [Sinohara H., 1998] have shown that DNA-hydrolyzing Bence Jones proteins from urine of patients with multiple myeloma are taken up by the cell and enter the nucleus, where they cause DNA fragmentation. The protein's activity was related to the progressive deterioration of clinical status, and catalytically active protein preparations were significantly cytotoxic. In addition, Kozyr et al [5] have shown the cytotoxicity of DNA-hydrolyzing SLE IgGs. Taking into account the above data, it is reasonable to suppose that even "non-specific catalytic Abs" from patients with different autoimmune diseases may be cytotoxic.

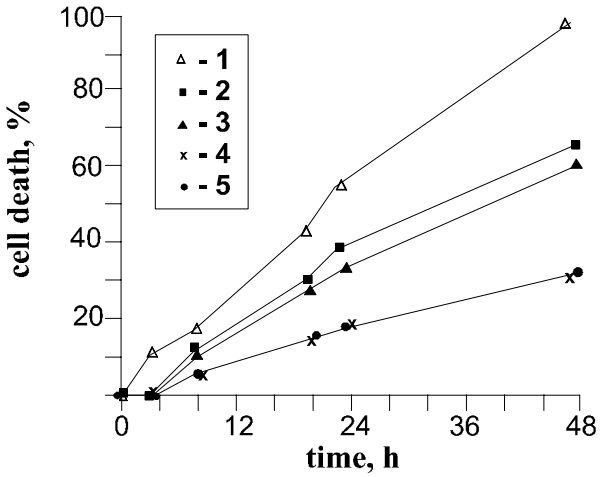


Figure 1. Time-dependent pattern of cytotoxicity of TNF-α (1); IgM SLE-abzymes (2); IgG MS-abzymes: patient 1 (3) and patient 2 (4), see text. Line 5 corresponds natural cell death (with medium alone).

Here we have compared for the first time a relative cytotoxicity of DNA-hydrolyzing Abs from the sera of patients with SLE and MS (Fig. 1). We have confirmed a previously reported finding concerning cytotoxicity of SLE DNA-hydrolyzing IgG [5] and have shown for the first time that DNA-hydrolyzing IgM of SLE patients may be also cytotoxic (Fig. 1, line *2*). It should be mentioned that the level of DNA- and RNA hydrolyzing activity of IgGs is strongly dependent on the patient, but in general increased in the order: diabetes ≤ viral hepatitis < polyarthritis ≤ Hashimoto's thyroiditis < SLE ≤ multiple sclerosis [3-5]. DNAase and RNAase activities of IgGs of MS patients usually comparable or even higher than those of patients with SLE. In spite of it, only one of 7 preparations of DNA-hydrolyzing IgG from MS patients (Fig. 1, lines *3* and *4*, respectively) possessed pronounced cytotoxicity comparable with that for SLE antibodies. It should be mentioned that we revealed RNA-, DNA-hydrolyzing IgG and/or IgM in the sera of patients with several autoimmune and viral pathologies [1-4], but we did not detect DNAase or RNAase activities of Abs in sera of 50 healthy humans. Interestingly, practically all BALB/c mice used in our experiments demonstrated detectable level of DNA-hydrolyzing activity. A possible reason of this phenomenon is not known now. This may be a common characteristic of BALB/c mice, or only a group of mice used.

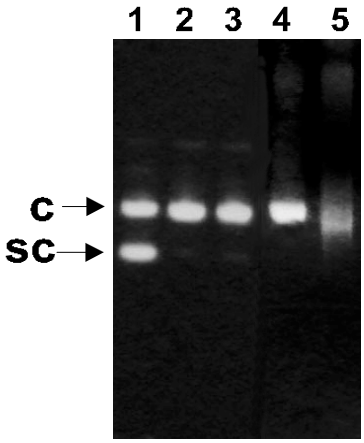


Figure 2. Induction of the DNAase abzymes in mice immunized with MS-IgG (3) and SLE-IgM (5). Lines 2 and 4 – controls of mice Abs before immunization, respectively; line 1 – control DNA plasmid pUC-18.

Immunization of BALB/c mice with DNA-hydrolyzing IgG and IgM antibodies from SLE patients led to a detectable increase of DNAase activity of Abs (Fig. 2) in the sera blood of mice (4 months after immunization). According to our previous data effect of MS abzymes is significantly lower than those of SLE Abs and sometimes it is absent at all. More detailed analysis of this question requires in future the statistical evaluation of the data. But even now one can see a correlation of a relative cytotoxicity and of relative effects of induction of catalytic antibodies by DNA-hydrolyzing antibodies from SLE and MS patients.

In order to analyze possible reasons for the differences in the effects of SLE and MS abzymes, the following should be taken into account. Recently we have shown that the substrate specificity of catalytic IgG isolated from sera of patients with different autoimmune and viral pathologies for classic poly(N) homopolynucleotide substrates and for specific $tRNA^{Phe}$ with compact and stable structure was correlated with the type of disease; the substrate specificity of major subfraction of polyclonal abzymes in the case of each disease is different in comparison with that for other pathologies. It means that in the case of each disease there may be different ways and/or different antigenes determining a production of catalytic Abs.

The complexity of the situation considered above is connected with several reasons. On one hand, abzymes may be Ab to analogs of transition states of catalytic reactions or even to substrates of enzymes playing a role of haptens. For example, cleaving vasoactive neuropeptide are Ab directly to the peptide. On the other hand, in the blood of patients with autoimmune diseases a spontaneous induction of antiidiotypic Ab to nucleic acids, proteins and their complexes is observed. In this context abzymes in autoimmune diseases are supposed to be Ab of antiidiotypic nature; they may be secondary Ab to the active centers of enzymes and contain idiotypes

giving "internal image" of the active center. Thus, in the case of different pathologies hypothetically there may be several different ways of formation of polyclonal abzymes with markedly different properties; Ab can be produced to the active centers of free enzymes, to their complexes with DNA or RNA, or directly to DNA or RNA in the their complexes with proteins. In the latter case the molecules of DNA and RNA can change their structure to a conformation modeling transition reaction state of nucleic acids in the reactions of their hydrolysis.

In this context it should me noted that in contrast to SLE abzymes, which are known as Ab predominately to double-stranded DNA, DNA-hydrolyzing MS-Abs may be mainly single-stranded DNA-dependent abzymes, since in contrast to SLE-Abs, MS immunoglobulins are more active in the hydrolysis of model single-stranded deoxyribooligonucleotides.

Taken together, it is obvious that in spite of an apparent similarity of DNA- and RNA-hydrolyzing Abs from the sera of patients with different autoimmune and viral diseases, these abzymes may be quite different in their role in pathogenesis of the above mentioned pathologies. The data obtained lead to many new questions concerning such DNAase and RNAase activities of IgGs and IgMs. Why do humans have such activities in their organisms? Why there are so many types of activities? Why repertories of abzymes are change very much from patient to patient and from one to other disease? Is it good or bad? A clear understanding will require further detailed investigations of possible similarities and differences of abzyme-dependent catalysis and a possible biological role of catalytic antibodies.

Acknowledgment

This work was supported by grants from the Russian Foundation for Basic Research (No. 98-04-49719), Russian Ministry of General and Professional Education ("Studies in Chemical Technologies", No. 98-8-3.2-69) and grants for young scientists from the Siberian Branch of Russian Academy of Sciences.

References

1. Vlassov, A. V., Baranovskii, A. G., Kanyshkova, T. G., Prints, A. V., Zabara, V. G., Naumov, V. A., Breusov, A. A., Giege, R., Buneva, V. N., Nevinsky G. A. (1998) Substrate specificity of serum DNA- and RNA-hydrolyzing antibodies of patients with polyarthritis and autoimmune thyroiditis. *Mol. Biol. (Moscow)*, **32**, 559-569.
2. Baranovskii A.G., Kanyshkova T.G., Mogilnitskii1 A.S., Naumov V.A., Buneva V.N., Gusev E.I., Boiko A.N., Zargarova T.A., Favorova O.O., Nevinsky G.A. (1998) Polyclonal antibodies from blood and cerebrosspinal fluid of patients with multiple sclerosis effectively hydrolyze RNA and DNA. *Biochemistry (Moscow)*, **63**, 1459-1469.
3. Andrievskaya O.A., Buneva V.N., Naumov V.A., Nevinsky G.A. (2000) Catalytic heterogeneity of polyclonal RNA-hydrolyzing IgM from sera of patients with lupus erythematosus. *Blood*, in press.
4. Vlassov A.V., Helm M., Florentz C., Naumov V.A., Breusov A.A., Buneva V.N., Giege R., Nevinsky G.A. (1999) Variability of substrate specificity of serum antibodies obtained from patients with different autoimmune and viral diseases in reaction of tRNA hydrolysis. *Russian J. Immunol*.,**4**, 25-32.
5. Kozyr, A.V., Kolesnikov, A.V., Aleksandrova, E.S., Sashchenko, L.P., Gnuchev, N.V., Favorov, P.V., Kotelnikov, M.A., Iakhnina, E.I., Astsaturov, I.A., Prokaeva, T.B., Alekberova, Z.S., Suchkov, S.V., Gabibov, A.G. (1998) Novel functional activities of anti-DNA autoantibodies from sera of patients with lymphoproliferative and autoimmune diseases. *Appl. Biochem. Biotechnol*., **75**, 45-61.
6. Paul S, Volle DJ, Beach CM, Johnson DR, Powell MJ, Massey RJ. (1989) Catalytic hydrolysis of vasoactive intestinal peptide by human autoantibody. Science, **244**, 1158-62.

PHYSICO-CHEMICAL CHARACTERISTICS OF THE FORMATION OF RecA-DNA NUCLEOFILAMENT

Bugreeva[1] I. P., Bugreev[1] D. V., Sinitsyna[2] O. I., *Nevinsky[1] G. A.

[1]Institute of Bioorganic Chemistry SB RAS, Novosibirsk, Russia
[2]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: Nevinsky@niboch.nsc.ru
*Corresponding author

Keywords: E.coli recA protein, filamentation onto single stranded DNA

Resume

Motivation:

By polymerizing on ssDNA RecA protein forms a right helical nucleoprotein filament (RecA-F), which can be several kilobases in length that provide DNA strand exchange reaction with the homologous dsDNA. The mechanisms that allow to realize such homologous alignment remain unclear and are one of the mysterious property of RecA protein. The principal effort to explain such homology recognition is guided by interaction between nucleotides strands within the filament. Nevertheless nuclear-protein interaction and the conformational changes occur during DNA recognition should be crucial for homology alignment. Thus the butter understanding of the thermodynamical characteristics of RecA-F formation may be important for revelation of function mechanisms of repair system.

Results:

Interaction of the RecA protein with single-stranded oligodeoxyribonucleotides (ssODNs) was investigated to determined the main factors of the formation of a RecA-ssDNA helical filament. In spite of the RecA can not form a stable nucleoprotein filament with short $d(pN)_n$ ($n \leq 16$), the protein can bind even dNMP. The results suggested that in filament RecA monomers forms the contacts with all $d(pT)_n$ monomers in an additive manner at least up to n= 20. The change in Gibbs energy due to addition of one nucleotide was evaluated to be~ 0.4 kcal/mole, that mainly reflects electrostatic contacts of RecA with ODN's sugar-phosphate backbone. Different modifications of ODNs such as ethylation of phosphate groups or removing of the bases significantly decrease the affinity but has not dramatic effect.

Introduction

Homologous genetic recombination is a central process in all-living organisms essential both for the generation of genetic diversity and for the repair of damaged DNA. The RecA protein is most crucial for the above processes: it identifies homology between two DNA molecules and catalyses the exchange of their strands. The reaction goes through three phases: (1) *presynapsis*, when RecA protein polymerizes onto ssDNA forming a right-handed helical nucleoprotein filament; (2) *synapsys*, during which dsDNA is bound by the presynaptic complex and brought to homologous alignment with the ssDNA strand; (3) *strand exchange*: the two like strands are exchanged producing a heteroduplex and a displaced linear single-strand.

In the presence of ATP RecA filament possesses three sites for the interaction with DNA. It can bind one molecule of ssDNA (or dsDNA and then ssDNA) and then coordinate one dsDNA or two additional ssDNA. The RecA-ssDNA-ATP complex readily binds dsDNA forming RecA-ssDNA-dsDNA-ATP complex, which is important for search of homology during synapsis and the following chain exchange.

In this paper we investigate the thermodynamical properties of RecA-ssDNA-ATP complex using the method of step by step simplification of the ligand structure that was previously proposed for the investigation of protein-nucleic acid interaction (for review see [1-2]). The possible mechanism by which a sequence similarity between two DNA molecules is recognized also has been discussed.

Methods

The RecA filament with ssODNs was formed at 30°C using $d(pT)_{20}$ as substrate. The reaction mixture (10 μl) contained 50 mM Tris-HCl 7.5, 10mM $MgCl_2$, 2mM DTT, 1mM γSATP, 0.1μM 5'-[^{32}P]$d(pT)_{20}$ and 1μM RecA. $d(pT)_n$ and different analogs were used as inhibitors of the filamentation reaction. Samples were incubated for 3 min with the following separation of product by retardation gel electrophoresis in TBE buffer. Gels were autoradiographed; the gel pieces corresponding to the products of free ODN and ODN in complex with RecA were cut out and counted for radioactivity. The affinity of ODNs was evaluated as I_{50} values (the concentration giving 50% inhibition of the activity of RecA-F formation).

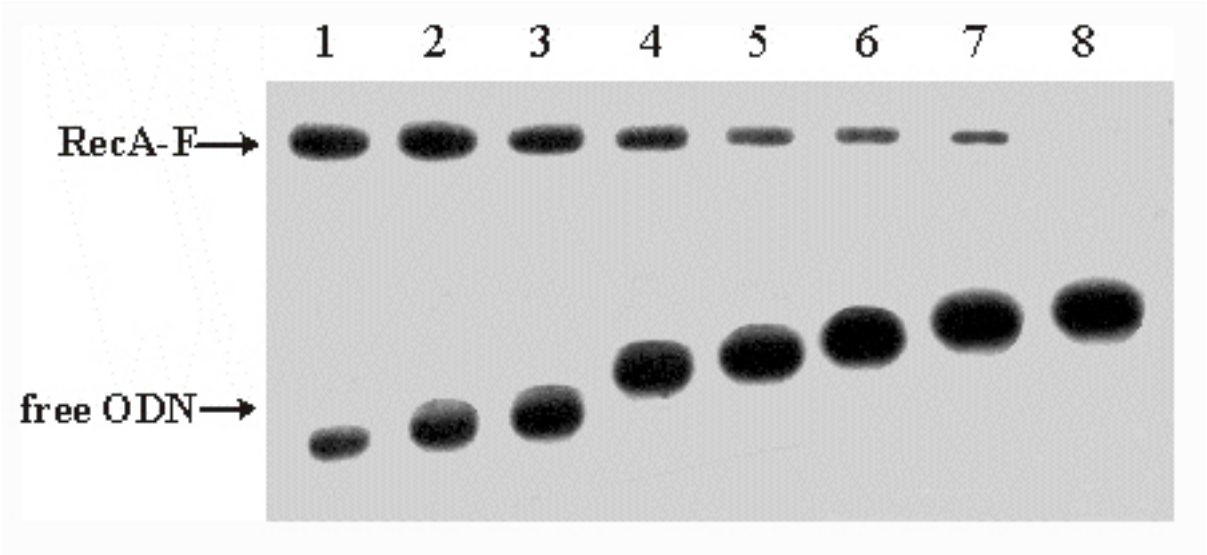


Figure 1. Inhibition of RecA–[32P]d(pT)20 filament formation by d(pT)16 oligonucleotide. The concentration of oligonucleotide used: 0.5 (line2), 1 (3), 2 (4), 4 (5), 8 (6) and 11 μM (7). Filament without inhibitor (line 1) and free oligonucleotide (line 8) were used as controls.

Results and discussion

As a rule, weak electrostatic, hydrophobic and van der Waals interactions do not play any significant role in the small ligand recognition by enzymes. However, this is not true for lengthy-DNA-dependent enzymes. We have recently developed several new approaches allowing to estimate a relative contribution of individual nucleotide units of ss and dsDNA to the total affinities of enzymes for long DNA. Detailed kinetic and thermodynamic investigations of pro-, eucaryotic and viral DNA polymerases, several repair, topoisomerization and restriction enzymes led us to conclusion that all long-DNA-dependent enzymes interact with all nucleotide units which are within their DNA binding clefts. For the 10-20 -unit-long DNAs the contribution of weak electrostatic, hydrophobic and van der Waals interactions to the total substrate affinity may be close to 4-8 orders of magnitude and predominate over the stronger specific and/or nonspecific ones due to the additivity of free energies of individual nucleotide units recognition. A new concept concerning a very important role of additive weak interactions providing high affinity of enzymes for DNA and specificity of their action was developed [1-2].

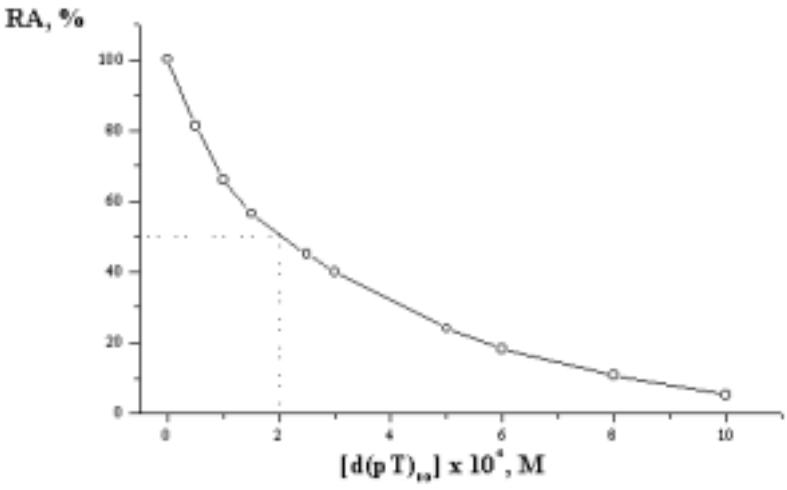


Figure 2. Concentration dependence of RecA remain activity (RA) by d(pT)10 oligonucleotide.

For all investigated enzymes, specificity of their action depends on the possible transformation of the associated DNA to its optimal conformation. The DNA adaptation process is very specific for each enzyme and proceeds more easily for specific rather than for nonspecific DNA. As a result, in the case of specific DNA k_{cat} increases of 3-7 orders of magnitude are found. A high level of specificity of catalysis in lengthy-DNA-dependent enzymes is provided by the catalytic constant as often happens in enzymology [1-2].

As was mentioned above the first step of homologous recombination involves the polymerization of RecA onto a ssDNA substrate in the presence of ATP, producing a filamentous nucleoprotein complex. This step is determinative for the overall process, because the initial binding which define structure and properties is likely to provide the following binding of dsDNA and homology recognition between DNA molecules.

Here we investigate the physico-chemical aspects of the RecA-ssDNA nucleoprotein filament formation. It was shown that filamentation of RecA protein onto 5'-[^{32}P]d(pT)$_{20}$ leads to formation of the filament, which is stable under conditions of electrophoresis (Fig. 1). The complexes of RecA with shorter ODNs are not stable during the gel retardation experiments. At the same time both shorter and longer ODNs were shown to be effective inhibitors of the RecA protein filamentation onto -[^{32}P]d(pT)$_{20}$ (Fig. 1).

Therefore for estimation of a relative affinity of different ODNs for RecA protein the I_{50} showing the concentrations of the ligand giving 50 % inhibition of the RecA-d(pT)$_{20}$ complex formation was used (for example, Fig. 2). The data obtained are presented as a logarithmic dependence of I_{50} values upon the d(pN)$_n$ length (n) (Fig. 3). As for all the above mentioned enzymes, the minimal ligand of RecA are orthophosphate (Pi, I_{50}= 0.53 M) and deoxyribose-5'-phosphate (I_{50}= 0,55 M). Transition from minimal ligands to d(pT)$_n$ (n=1-20) leads to gradual increase of ODNs affinity. The linearity of the dependence (Fig. 3) provides evidence for the additivity of the free energies characterizing interaction of the individual d(pT)$_n$ units with RecA monomers of the filament.

It is known that the RecA-F can be several kilobases in length but each RecA monomer in the filament interacts with 3 nucleotide units of ss or ds DNA. Taking this into account from data of Fig. 3 one can see that each RecA monomer interact with all 3 nucleotide units of ODNs with nearly the same efficiency. Since RecA-F can be very large in size, it is very interesting weather all RecA monomers of the filament of several kilobases can interact with ss DNA in an additive manner. As one can see from Fig. 3, relative affinity of d(pT)$_{30}$ (I_{50}=2.3•10^{-8} M) and d(pT)$_{40}$ (I_{50}=7.0•10^{-9} M) significantly higher than that for d(pT)$_{20}$. However, the lgI_{50} values for these ODNs do not correspond to the linear of for curve for (pT)$_1$ –d(pT)$_{20}$.

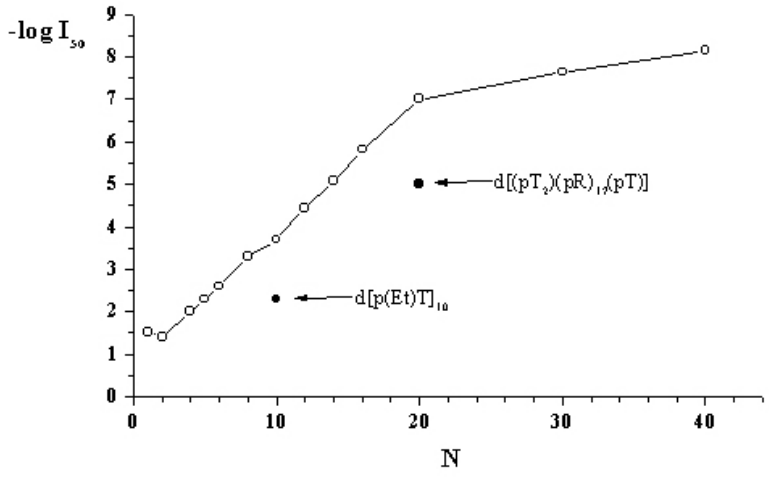


Figure 3. Dependence of logI50 on oligodeoxythymidilate length (N).

This phenomenon may be a consequence of two different reasons. Firstly, the protein concentration, which required for filamentation of RecA protein onto d(pT)$_{20}$ is significantly lover than that for the RecA complex formation on d(pT)$_{30\text{-}40}$. In addition, filamentation of RecA protein on very long ssDNA is a time consuming process and it can not be described using initial rates of the filamentation reaction. At the same time linear decrease of the affinity of ODNs for d(pT)$_1$ –d(pT)$_{20}$ and then for d(pT)$_{20}$- d(pT)$_{40}$ can indicate for similar additive contacts of RecA monomers with all nucleotides of ss DNA of any length.

From the slope of the curve (Fig. 3) the incremental factor for ODN lengthening by one nucleotide was evaluated to be equal to ≈2.1, which corresponds ΔG^o = – 0.4 kcal/mole. This value is essentially lower than the free Gibbs energy characterizing formation of enzyme's strong contacts with small ligands, but is a comparable with that for weak additive contacts between long DNA and many DNA-dependent enzymes [1-2].

Interestingly, transition from d(pT)$_{10}$ to its ethylated analogue d[p(Et)T]$_{10}$ decreases the affinity by a factor of ~25. In addition, the affinity of d(pT)$_8$ (I_{50}=5•10^{-4} M) containing 5' phosphate is about 50-100 times higher than that for d[T(pT)$_7$] (I_{50}=4.8•10^{-3} M) and for d(Tp)$_8$ (I_{50}=2.5•10^{-3} M). All this data indicate for important role of negative charges of internucleoside phosphate groups of ODNs for their interaction with RecA filament. At the same time, the relative affinity of d[(pT$_2$)(pR)$_{17}$(pT)], containing 17 abasic units (R) is about 2 order of magnitude lower than that for d(pT)$_{20}$. Consequently, weak additive interactions of RecA monomer with monomers of ssDNA are a combination of the protein interaction with both bases and internucleoside phosphate groups of ssDNA. A more detailed study of a relative contribution of sugar-phosphate backbone and different bases of DNA to the total affinity of RecA filament for ss DNA is in a progress.

References

1. Nevinskii, G.A. (1995) An important role for weak interactions during the recognition of long DNA and RNA molecules by enzymes. *Mol. Biol. (Moscow)* **29**, 16-37.
2. Bugreev D. V, Nevinsky G. A. (1999) Possibilities of the method of step-by-step complication of ligand structure in studies of protein--nucleic acid interactions: mechanisms of functioning of some replication, repair, topoisomerization, and restriction enzymes. *Biochemistry (Moscow)* 64, 237-249.

INFORMATIONAL SUPPORT OF ONTOLOGY FOR TRANSCRIPTION REGULATION (ISOTR)

Ignatieva E.V., Podkolodny N.L., Kolchanov N.A.

Institute of Cytology and Genetics of SB RAS, Novosiborsk, Russia
e-mail: eignat@bionet.nsc.ru
*Corresponding author

Keywords: informational computer resources, gene expression regulation, ontology, meta-data, vocabulary

Resume

Motivation:

A successful theoretical analysis of all aspects of live systems and objects functioning is possible only in case there exists a considerable bulk of data and knowledge on these systems and objects. Due to rapid development of novel methods and approaches aimed at studying gene expression regulation, contemporary molecular biology was enriched by huge massive of information on cell structures and regulatory mechanisms of gene transcription. For effective involvement of specialists in various disciplines to theoretical analysis of gene expression regulation at the level of transcription, it is very useful to create an informative resource that will provide an access to the most up-date knowledge on this problem. This resource should obligatory include the system of basic concepts on gene transcription regulation. Besides, it should provide a possibility of effective and complete extraction of information, concerning the problem of interest, from MEDLINE database.

Results:

In the present work, we present a project of informational support of ontology for transcription regulation (ISOTR). The system will include hierarchically organized and controlled curated vocabularies and the database of annotated MEDLINE entries. Two regimes of the database accumulation are allowed, namely, the manually operated regime and the regime supporting automated queries to MEDLINE.

Introduction

A topical problem of bioinformatics is development of computer-assisted methods for analysis of genome sequences in order to reveal the regions with different properties, i.e., transcribed sequences, regulatory regions (promoters, enhancers, LCR, MAR, etc.). Furthermore, the subsequent analysis of functional activity of regulatory regions will be done. Contemporary computer technologies designed for analysis of structure-functional genome organization are based on application of databases and knowledge base, which enable to compare definite properties and parameters of the sequences analyzed with the results of analogous analysis performed for the learning samples. Thus, the presence and availability of appropriate computer information resources (databases and knowledge bases) are mainly determined by possibilities of computer-assisted methods of analysis.

Transcription is a key event initiating the whole integrity of processes necessary for gene expression regulation. This fact explains a great interest to the studying of mechanisms controlling gene expression at the level of transcription in eukaryotes. During analysis of gene expression regulation at the level of transcription, such informational resources as TRRD [Kolchanov,N.A. et al., 2000a], ACTIVITY [Ponomarenko,M.P. et al., 1999], TRANSFAC [Wingender,E. et al., 2000], SELEX [Ponomarenko,J.V. et al., 2000] are widely used. The databases and knowledge bases mentioned above are characterized by the stable format with the strictly fixed type of the data introduced into these databases. In addition, intense development of modern methods designed for the study of gene expression regulation [Kadonaga,J.T., 1999; Simpson,R.T, 1999] enabled to receive the data indicating to a huge variety of the mechanisms on transcription regulation, of the types of regulatory genome sequences, spectrum of protein factors involved into this process, and the ways of interaction between these factors. These novel experimental data are often can not be described within the frames of the existing database formats. However, this information may be of extreme importance both for operative tracing for dynamics of accumulation of scientific facts and their variability and for decision-making on modification of the formats of the databases on gene expression regulation. Thus, due to necessity of the further development of specialized databases on gene transcription regulation, we believe that accumulation and systematization of this data flow are very important. Due to above reasoning, we set the task to form a concept

of a flexible information resource that will fulfil the role of the catalogue of structures, processes, and mechanisms involved in the gene transcription regulation. It is reasonable to design such information resource on the basis of ontological description of the subject environment. As an ontology, we understand "the system describing concepts and the relationships between them" [Barker,P.G., 1999]. Currently, at the Institute of Cytology and Genetics, a work is in process, which is devoted to development of ontology on gene expression regulation and to development of databases based on annotating of scientific publications [Kolchanov,N.A. et al., 2000 (a); Kolchanov,N.A. et al., 2000 (b)]. An information resource under discussion may perform the informational support of this process. The hierarchically organized, controlled, and curated vocabularies will be a core of the suggested information resource ISOTR (Information Support of Ontology for Transcription Regulation). Besides, this system will include the tools installing the logic relations between the terms from vocabularies. Additionally to vocabularies, the system ISOTR will contain the MEDLINE entries with information referring to the terms from vocabulary. At last, a discriminative feature of the system suggested is a possibility of its accumulation performed in two regimes, automated and manual ones (Fig.1).

The structure of the system "ISOTR"

Annotated entries from MEDLINE. The annotated entries will be created on the basis of MEDLINE entry in the format "MEDLINE report". It is suggested that there will be annotated entries of two types: 1) entries of experimental papers; and 2) entries of review articles (Fig. 2). The type of an entry will be fixed in a special field. In the course of annotation, the entry corresponding to experimental paper will be supplemented by information on the gene name, its species-specificity, and keywords from the vocabularies. This information will be registered in a formalized form in the fields of respective types (Fig. 2). Except the data mentioned above, the annotated entry may contain the information on the types of cells and tissues, inductors, stages of development, and the other conditions essential for the given study. If the initial MEDLINE entry has no information of the certain type, the corresponding field stays empty. The entries corresponding to review articles will be supplemented mainly with the keywords.

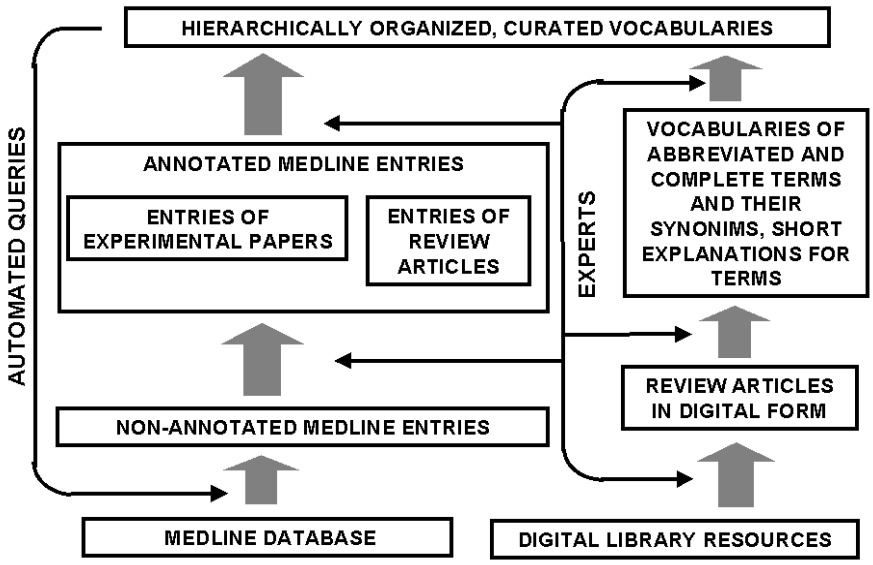


Figure 1. The structure and accumulation of "ISOTR"

Annotated entries will be cross-linked with the vocabularies from the system ISOTR, along with the table of genes (TRRDGENE) and vocabularies of cells, organs, tissues, and inductors from the TRRD database [Kolchanov,N.A. et al., 2000a].

Not-annotated entries from MEDLINE. It seems reasonable to store not-annotated entries from MEDLINE that were extracted by automated queries to MEDLINE. Moreover, the entry will be automatically supplemented by the keyword, in accordance to which the query was made.

Hierarchically organized curated vocabularies. As the central elements of the ISOTR stem will serve the vocabularies including the names of structures and processes related to transcription regulation. Besides, the catalogs including the models and suppositions, together with detailly discussed mechanisms of functioning, observations, and other data on the features of gene expression will be organized in a form of vocabularies. The vocabularies will have hierarchical structure (Fig. 3) and the experts will fill them in by annotating MEDLINE entries or review articles.

Other types of data. Except hierarchically organized vocabularies, it is possible to support the vocabularies of abbreviated and complete names of terms (Fig. 1). Accumulation of these vocabularies will be made with application of keywords from review articles published in the journals like "Current opinion" and so on. For a seria of keywords, corresponding to the basic concepts, the brief explanations will be done. Possibly, an information module will be constructed that will contain modern reviews on transcription regulation.

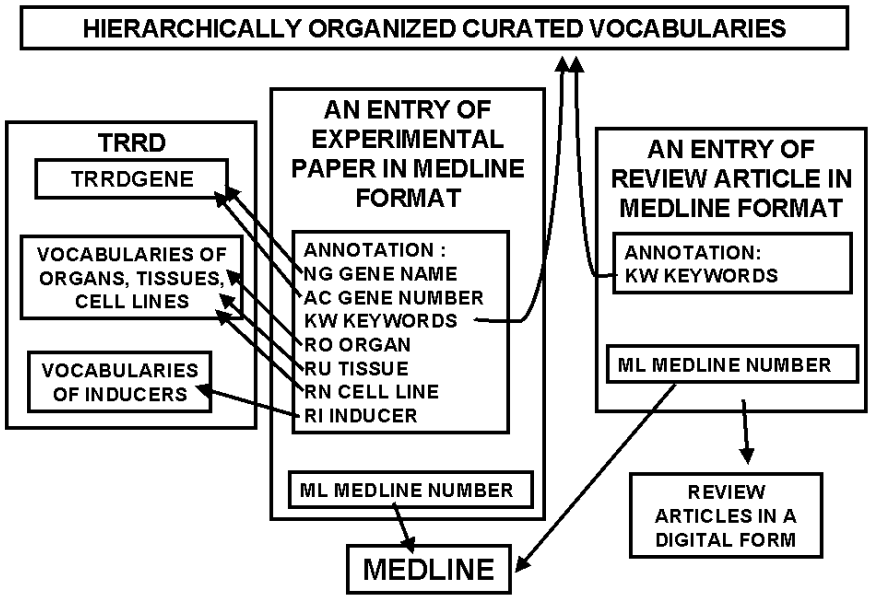


Figure 2. Format of the annotated MedLine entry and references to the other information resources.

Inclusion of the new data in the system "ISOTR"

Manual regime. The experts will do annotation of MEDLINE entries corresponding to review and experimental papers. In the course of annotation, vocabularies will be supplemented with new keywords.

Automated regime. Extraction of entries without annotations will be done on the basis of automated queries to MEDLINE by using keywords from the vocabularies of the system ISOTR. The usage of combinations of keywords is also possible. The entries extracted in such a manner will be stored with adding of information on the keywords that were used for its extraction.

Conclusion

In the present work, we introduce a project ISOTR that will accumulate the contemporary concepts on mechanisms and structures involved into transcription regulation. The presence of two systems for ISOTR accumulation (i.e., manual and via automated queries) determines rather vide range of its application. For instance, the system of keywords supported within the frames of ISOTR system fulfills simultaneously two functions. The first is the usage of the terms from vocabulary as the keywords for the MEDLINE-assisted search. This is a very effective way to extract information of interest. Second, the vocabularies may be used as catalogue of modern knowledge on the hot problem. An appearance of novel keywords in the catalogue is a signal that there is an information of the necessary type in information resource given. Thus, a possibility appears to evaluate novel tendencies in accumulation of information, which, in its turn, will be a basis for decision-making on development of novel databases.

Acknowledgements

This work is supported by the Russian Foundation for Basic Research (grants Nos 98-04-49479, 98-07-91078, 98-07-90126, 99-07-90203, 00-04-49229, 00-04-49255, 00-07-90337) and the grant by US Department of Energy (DE-FG02-00ER62893/535228). The authors are grateful to O.A. Podkolodnaya for fruitful discussions and to I.V. Lokhova for the bibliographic assistance.

References

1. Baker,P.G., Goble,C.A., Bechhofer,S., Paton,N.W., Stevens,R., Brass,A. (1999) An ontology for bioinformatics applications. *Bioinformatics* 15, 510-520.
2. Kadonaga,J.T., Grunstein,M. (1999) Chromosomes and expression mechanisms Chromatin: the packaging is the message. *Current Opinion in Genetics & Development,* 9, 129-131.
3. Kolchanov,N.A., Podkolodnaya,O.A., Ananko,E.A., Ignatieva,E.V., Stepanenko,I.L., Kel-Margoulis,O.V., Kel,A.E., Merkulova,T.I., Goryachkovskaya,T.N., Busygina,T.V., Kolpakov,F.A., Podkolodny,N.L., Naumochkin,A.N., Korostishevskaya,I.M., Romashchenko,A.G., Overton,G.C. (2000a) Transcription Regulatory Regions Database (TRRD): its status in 2000. *Nucleic Acids Res.,* , 28, 1, p. 298-301

4. Kolchanov N.A., Podkolodny N.L., Ponomarenko M.P. et al., (2000b) Integrated system on gene expression regulation – GeneExpress. This issue.
5. Ponomarenko,J.V., Orlova,G.V., Ponomarenko,M.P., Lavryushev,S.V., Frolov,A.S., Zybova,S.V., and Kolchanov,N.A. (2000) SELEX_DB: an activated database on selected randomized DNA/RNA sequences addressed to genomic sequence annotation. *Nucleic Acids Res.,* 28, 205-208.
6. Ponomarenko,M.P., Ponomarenko,J.V., Frolov,A.S., Podkolodny,N.L., Savinkova,L.K., Kolchanov,N.A., and Overton,G.C. (1999) Identification of sequence-dependent features correlating to activity of DNA sites interacting with proteins. *Bioinformatics,* 15, 687-703.
7. Simpson,R.T. (1999) In vivo methods to analyze chromatin structure. *Curr.Opin.Genet.Dev.* 9, 225-229.
8. Wingender,E., Chen,X., Hehl,R., Karas,H., Liebich,I., Matys,V., Meinhardt,T., Pruss,M., Reuter,I., Schacherer,F. (2000) TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.* 28, 316-319.

1. STRUCTURES VOCABULARY

1.1. Regulatory regions

1.1.1. Promoter*
1.1.2. Core promoter
1.1.2.1. Initiator
1.1.2.2. TATA box
1.1.2.3. TATA+ Inr+
1.1.2.4. TATA+ Inr-
1.1.2.5. TATA- Inr+
1.1.2.6. TATA- Inr-
1.1.2.7. RNA-polymerase-I-transcribed promoter
1.1.2.8. RNA-polymerase-II-transcribed promoter
1.1.2.9. RNA-polymerase-III-transcribed promoter
1.1.2.10. distinct core promoters
1.1.2. Promoter proximal elements*
1.1.3. Enhancers *
1.1.4. Silencers*
1.1.5 Repressor elements*
1.1.5. Boundary chromatin elements*
1.1.6. DNase I hypersensitive sites*
1.1.7. Four-way junction DNA

1.2. Multiprotein complexes and proteins, regulating transcription

1.2.1. Preinitiation Complex (PIC)*
1.2.2. Holoenzyme*
1.2.5. Coactivators*
1.2.6. Corepressors*
1.2.7. Chromatin remodeling machines*
1.2.8. Histone acetyltransferases and deacetylases*
1.2.9. Protein components of SARs, MARs, and insulator elements*

1.3. Chromatin*
1.4. Nucleosome*
1.5. Transcription factories*

2. EVENTS AND PROCESS VOCABULARY

2.0. Transcription cycle

2.0.1. Preinitiation complex formation
2.0.2. Open complex formation
2.0.3. Promoter escape
2.0.5. Termination
2.0.6. Reinitiation

2.1. Transcription initiation*
2.2. DNA-protein interactions*
2.3. Protein-protein interactions

2.3.1. Protein-protein interactions between components of PIC*
2.3.2. Protein-protein interactions between PIC and transcription factors or coactivators
2.3.3. Protein-protein interactions between transcription factors and coactivators

2.4. Wrapping of DNA*
2.6. Histone acetylation *
2.7. DNA methylation*
2.8. Chromatin remodelling*
2.9 Phosphorylation*

3. MECHANISMS VOCABULARY

3.1. The mechanism for transcriptional synergy by eukariotic activators
3.2. The mechanism for transcription activation by torsional transmission
3.3. The mechanism of alternative regulation of RNA polymerase II and RNA polymerase III
3.4. Nucleosome positioning
3.5. Nucleosome and transcription factors competition

Figure 3. Fragments of hierarchically organized vocabularies from ISOTR. Asterisks mark sections, which contents are not shown in the figure.

METHOD OF FUNCTIONAL SITES CLASSIFICATION BASED ON OLIGONUCLEOTIDE FREQUENCIES

Nedosekina E.A., *Pozdnyakov M.A.

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
e-mail: mike@bionet.nsc.ru
*Corresponding author

Keywords: classification, functional sites, oligonucleotide frequencies, tissue specificity, discriminant analysis

Introduction

For functional site recognition in nucleotide sequences, different methods are used such as consensi, weight matrices [1,2,3,4,5,6], neural networks, discriminant analysis [7,8,9] etc. However, despite of the durable research, the problem of exact functional site recognition is still unsolved. We suggest that one of the reasons explaining this fact could be the heterogeneity of functional site sequence sets, which are used for construction of recognition methods. Such a heterogeneity could produce several groups in the set of sequences. These groups differ by the context features, although the sequences are characterized by the similar molecular function. Such a redundancy of context features respective to function may be caused by the functional specificity. Therefore, site activity is defined mainly by conformational properties of the site, which could be similar for the sites with significantly different nucleotide sequences. In this case, evolution process maintains only conservation of certain conformational properties of functional sites, whereas their primary structures may diverge. The second reason is explained by differential site functioning in various tissues or under different physiological conditions. Regulatory proteins and protein complexes interacting with the sites of the same type may have tissue specific differences. Taking these facts into account, we suggest that at the first step of any site recognition method design, classification of the site set should be produced into the groups homogeneous by the context. In this work, we suggest the method aimed at classification of the sites based on discriminant analysis of oligonucleotide frequencies. The method was applied for classification of donor splicing sites of human tissue specific genes.

Methods

We have developed the method of sites classification based on discriminant analysis of oligonucleotide frequencies. Input data of this method are (1) the length of oligonucleotides L; (2) the ratio of false-negative and false-positive errors Q; (3) the number of regions R, in which all the sites are divided so that oligonucleotide frequencies could be defined separately in each of such regions; (4) sets of true and false sites of equal length.

The results of the algorithm processing are as follows: (1) a set of classes, in which the initial sets of true and false sites are divided; (2) mean oligonucleotide frequencies for each class; (3) the set of linear discriminant functions for each pair of classes.

At the beginning, there are two classes: one contains all true and another - all false sites. Algorithm defines linear discriminant function, which uses oligonucleotide frequencies of length L to discriminate these two classes. It was proposed that the values of discriminant function follow normal distribution on the sets of true and false sites. The mean and standard deviation values were achieved for each class. By using them and the parameter Q, the following values were estimated: (1) the values of false negative and false positive errors; (2) the threshold value of discriminant function.

For simplicity, let's consider Q=1. Under condition that oligonucleotide frequencies of two classes follow the same distribution, randomly chosen discriminant function in average attributes half of sites of each class properly (to the same class) and erroneously attributes the second half to another class. The constructed discriminant function attributes properly and erroneously the other part of sites. The deviation of real predictions from the expected ones was estimated by the chi-square criterion $\chi 2$. The freedom degree numbers N equals to 1 for each pair of classes (initially there is only one pair). When the number of classes increases, the values $\chi 2$ and N are summed up for all pairs of classes. P-level for this summed $\chi 2$ and N can be used as the measure of quality of classification. In the method developed, we have used z-statistics of this p-level as the measure of classification quality, z(p). For obtaining z(p), the sets of random sequences were generated many times with the same nucleotide frequencies as in real sequences considered. For these sets, the p-level was obtained as described above. All generated random sets were stored for the subsequent class divisions and for obtaining z(p) at the next steps of the algorithm's work.

In what follows, the algorithm produces iteratively decisions which class should be divided into two. For this purpose, the algorithm (1) estimates z(p) in case the class of true sites is divided; (2) estimates z(p) in case the

class of false sites is divided; (3) chooses which z(p) is higher and makes the respective class division. During the algorithm's processing, the number of classes containing true and false sites increases. For each pair of classes, the linear discriminant function and the portions of erroneously attributed sites for each pair of classes are obtained. Only a single class is divided into two during one iteration. When there appear several classes of true sites, an algorithm chooses one of them according to the largest portion of erroneously attributed sites to any of false sites class. Analogously, the class of false sites is chosen. The decision, which class (i.e., of true or false sites) should be divided, is made by using the value z(p) for each case, as described above. For obtaining z(p), the stored random sets were used, and the choice of classes to be divided was conducted in the same manner.

The optimal number of classes is unknown beforehand, hence, it is chosen by the algorithm like that for choosing the highest value of z(p). So, the algorithm can divide the set of true sites, the set of false sites, both of them or choose the initial situation with undivided set as the best one.

Results

We have analyzed the sets of donor splice sites of human tissue specific genes in lung (38 sites), liver (409 sites), kidney (457 sites), and brain (20 sites). Gene sequences were extracted from EMBL, release 62. Donor splice sites were considered together with the flanking regions of 20 bp in length. As the sets of false sites, the random sequences were generated with the same nucleotide frequencies as in true sites, with GT conservative dinucleotide at the exon/intron boundary. The parameters of the method were: Q=1, R=1, L=2. This means that each site was viewed as a single region where dinucleotide frequencies were considered. Only for the brain specific splice sites, a heterogeneity of splice sites was detected. Two classes of true splice sites, each containing 10 sites, and two classes of random sites were obtained. Z-statistics for classification resulted was 3.86, this value corresponds to the p-level 6×10^{-05}. The dinucleotide frequencies for each class are listed in Table 1.

Table 1. Absolute dinucleotide frequencies in classes of true and false donor splice sites of brain specific human genes.

	Mean dinucleotide frequencies															
	AA	**AC**	**AT**	**AG**	**CA**	**CC**	**CT**	**CG**	**TA**	**TC**	**TT**	**TG**	**GA**	**GC**	**GT**	**GG**
True, class 1	1,3	1,5	1,1	2,9	2,3	3,2	3,8	1,7	0,8	1,3	0,8	4,8	2,4	4,7	2,1	4,3
True, class 2	3	2	1,1	4,3	2,6	2,3	1,5	2,2	1,6	1,9	1,1	2,1	3,2	2,6	2,7	4,8
False, class 1	1	2	1	2,4	2	3,6	2,3	4,5	1	2,5	1,2	2,4	2,4	4,4	2,6	3,8
False, class 2	3,2	2	1,6	3,8	2,2	1,6	1,2	2,4	2	1,4	1,6	2,4	3,4	2,2	3	4,8

Under supposition that all dinucleotide frequencies are the same, the absolute dinucleotide frequencies in Table 1 should equal to 40/16 = 2,5 (40 – site length). As can be seen from Table 1, there are dinucleotides occurring with the frequencies strongly deviating from expected ones. For example, in the first class of true sites, TG, GC, and GG dinucleotides occur with increased frequencies, whereas the occurrence frequency of AA, AT, TA, TC, and TT is decreased. Besides, the differences were found in dinucleotide frequencies between true and false sites. For example, AA and CT dinucleotide frequencies in the first class of true sites equal to 1.3 and 3.8, respectively, whereas in the second class of false sites, they are 3.2 and 1.2, respectively. Classes of true and false sites differ from each other by the number of dinucleotide frequencies (for example, the frequencies of AA, CT, and TG dinucleotides). All in all, one can see that the differences in dinucleotide frequencies between true and false sites are complicated, that is why we have applied the method of discriminant analysis.

References

1. M.S. Gelfand, J.Comput.Biol, 2, 87-115, 1995.
2. Bucher, J.Mol.Biol.,212, 563-578, 1990.
3. Karlin and V.Brendel, Science, 257, 39-49, 1992.
4. K.Quandt, K.Frech, H.Karas, E.Wingender and T.Werner, Nucleic Acids Res., 23, 4878
5. E.C.Uberbacher, Y.Xu and R.J.Mural, Methods Enzymol., 266, 259-281, 1996.
6. Q.K.Chen, G.Z.Herz and G.D.Stormo, Comput Appl. Biosci, 13, 29-35, 1997.
7. V.V. Solovyev, A.A. Salamov and C.B. Lawrence, Nucleic Acids Research, Vol.22, 5156-5163, 1994.
8. V.V.Solovyev, A.A.Salamov and C.V.Lawrence, In Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. In Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology. AAAI Press, 367-375, 1995.
9. M.Zhang, Proc.Nat.Acad.Sci., 94, 565-568, 1997.

DROSOPHILA PROMOTER DATABASE EnDPD: PROJECT AND THE FIRST STEPS OF ITS REALIZATION

Katokhin A.V., Levitsky V.G.

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: katokhin@bionet.nsc.ru
*Corresponding author

Keywords: Drosophila, eukaryotic promoters, transcription start sites, transcription elements, database

Resume

Motivation:

Computer-assisted prediction of eukaryotic promoters is still one of the stubborn problems appearing in analysis of extended nucleotide sequences. The solving of this problem depends heavily on the way of organizing the sample of promoter sequences. One of the most important sources of structure-functional information on promoters became the EST and Genome projects. The database accumulating all available information on promoters of a single model organism, *Drosophila melanogaster*, gives broad analytical possibilities for studying fine organization and recognition of promoters of this species. Moreover, it is useful for working out a concept on the structure and function of promoters in eukaryotes.

Results:

EnDPD (Enhanced Drosophila Promoter Database) is realized as an advanced informational resource, which supplies a user with an ability to get information about localization, structure, and function of some promoter of *Drosophila melanogaster* gene by using hyperlinks. At the firs step, EnDPD unifies the data on promoters of 250 genes. Each promoter description is supplied by 1) hyperlink to FlyBase and GADFly entries related to the gene of interest; 2) coordinates of transcription start site (TSS) location (given previously in literature) in correspondence with the numbering of Celera/BDGP's genome contigs; 3) data on localization of 5'-ends of corresponding ESTs.

Introduction

In recent years, when the projects on genome sequencing of several model organisms are being successfully finished, development of the methods aimed at automated genome annotation became one of the most rapidly developing directions in genomics. The goal of annotation procedure is to extract maximum information out of extra-large nucleotide sequences in order to formulate the knowledge on gene location, structure, and the way of its functioning.

It is of common knowledge that the level of gene transcription regulation produced by interaction of protein transcription factors with promoters is the most significant for their differential expression. The large impact into development of computer methods aimed at analysis of eukaryotic promoter structure was made by the database EPD (Bucher,Trifonov,1986, Bucher,1997).The sample of promoters from EPD was used as the training set of almost all the programs on promoter recognition in genome nucleotide sequences.

However, many researchers should accept that up-to-date computer software packages are unable to predict reliably transcription regulatory element location in genes (Fickett,Hatzigeorgiou, 1997; Duret,Bucher,1997; Bucher,1999; Pedersen et al.,1999). This fact is mainly explained by insufficient accounting of variety of promoter structures, which reflects the variety of species-, tissue-, and stage-specificity conditions of their functioning, revealed by modern experimental data. Thus, the main goals in developing computer-assisted methods for promoter recognition are switching from analysis of general eukaryotic promoter structures to analysis of organization of particular promoter classes and sub-classes (Pedersen et al.,1999).

Drosophila melanogaster is one of the leading model organisms, because the data on complete genome structure of this object go advantageously with numerous and various data on phenotypic gene manifestation and gene interaction. This enables to apply Drosophila genome sequences for testing an accuracy and susceptibility of many methods developed for automated gene recognition (Reese et al,2000a, Salamov, Solovyev,2000, Krogh,2000). As a result of such testing within the frames of GASP (Genome Annotation Assessment Project), important conclusions were obtained on efficiency of the algorithms developed. Moreover, it was stated that promoter prediction in Drosophila is still a challenging task (Reese et al, 2000b). One of the reasons explaining this facts may be the evidence that there are definite difficulties in experimental detection of transcription start sites (TSS). With this respect, the main goal under development of this or that algorithms is

very careful compilation of promoter sequence set, by using all the variety of structure-functional information about promoters.

The foundation of the database represented is a set of Drosophila promoter sequences (DPD), the collection of sequence data on 252 promoters in *Drosophila melanogaster*,compiled by Arkhipova I.R. (Arkhipova, 1995). It is available at BDGP server (http://www.fruitfly.org/ sequence/ drosophila-datasets.html). Since initial development of DPD five years ago the value of this set aimed at analysis of promoter structures of a single species was successfully demonstrated both in context studies (Arkhipova,1995; Burke, Kadonaga,1996), and in the process of development of novel methods for promoter prediction (Ohler et al., 1999; Reese et al., 2000a; Levitsky et al., 2000). To increase an efficiency of these studies, it is necessary to take into account and to systemize the novel data produced by EST ang genome projects in Drosophila. In order to integrate the data from several sources, we have developed a curated database EnDPD. The structure and format of EnDPD correspond to contemporary requirements for informational resources. For example, each promoter description is supplied by hyperlinks: 1) to the other necessary information on corresponding genes; 2) on results of promoter analysis given in a graphical visualization; 3) on software programs for context analysis or comparative study of novel promoters.

Results

Description of EnDPD database. The principles of data representation in EnDPD are mainly correspond to those developed in the last releases of EPD (Perier et al.,1998; Perier et al.,1999) (http://www.epd.isb-sib.ch/). The significant features of EnDPD as a database for a single species are (i) the hyperlinks, besides those to FlyBase (http://flybase.bio.indiana.edu:7083/), to the other important databases on structure and functions of Drosophila genes: BGDP (http://www.fruitfly.org/) and GADFly (http://flybase.bio.indiana.edu:7083/annot/); (ii) the absence of references to the literature sources, because for access to initial literature data, it is sufficient to have a link either to FlyBase document (which , in its turn, gives an access to corresponding InteractiveFly document) or to EMBL/GenBank entry.

```
ID  Rbp9'2
XX
AC  DP0180
XX
DT  01.06.2000
XX
DE  23C1-2
XX
FR  FLYBASE; FBgn0010263; Rbp9.
FR  GADfly; CG3151:CT38165
XX
ER  EMBL; L04930:2288.
XX
CK  Celera/BDGP scaffold; AE003581:-179733.
CK  Celera/BDGP contig; AE002638:-1986180.
XX
AP  Alternative promoter #2 of 3; exon 2
XX
NP  none
XX
ES  BDGP EST; GM26440/AI406214 (32/1)
ES  BDGP EST clot#;.
XX
ME  Nuclease protection
XX
EX  Pupa; Adults: cell body of adult brain,
EX  thoracic ganglion, antenna, ovariole
XX
FG  http://wwwmgs.bionet.nsc.ru/Image/DP0180
XX
TL  http://wwwmgs.bionet.nsc.ru/Programs/
XX
SQ  ….
//
```

Figure 1. An example of entry in EnDPD database

All 88 promoters of *Drosophila melanogaster* described in EPD were also accompanied by novel data and then they were accumulated in EnDPD.

In Fig. 1, one of the entries documented in EnDPD is given, this entry describing RNA-binding protein 9 gene promoter.

Informational content of entries in EnDPD. An entry in EnDPD contains the following fields with information about the gene and its promoter:

- ID – name of promoter based on gene denotation in correspondence with FlyBase nomenclature,
- AC – accession number
- DT – date of creation
- DE – name of a gene and its most distributed synonyms
- FR – links to corresponding entries in FlyBase and BDF/BGDP/GADFly)
- ER – link to EMBL/GenBank entry containing initial information about TSS location
- CK – coordinates of TSS with respect to numeration of scaffold contigs (available for BLAST-analysis) and supercontigs Celera/BGDP
- AP – information about possible alternative promoters in a gene
- NP – information about adjacent promoters
- ES – reference to the EST most 5'-located, with indication of its position relatively TSS, and reference to a clot, clustering other ESTs of respective gene
- ME – description of experimental evidence supporting TSS detection
- EX – information on respective gene expression properties provided by the promotor

- FG – link to the profiles visualizing the results of context analysis of promoter nucleotide sequence
- TL – link to software programs enabling to carry out customized context analysis of promoter nucleotide sequence
- SQ – promoter nucleotide sequence encompassing –700 +300 bp region around TSS.

Discussion

The first task was to estimate one-to-one correspondence between promoters and concrete genes and denote them with respect to FlyBase classification. This enabled to fulfill an important criterion formulated in EPD, that is, non-repeatability of the objects in the database. In case multiple TSSs were documented in literature for a single exon of an arbitrary gene, only those TSS positions separated by at least 100 bp were related to different promoters.

The criterions of including a promoter into EnDPD database, basing on quality and quantity of experimental data, are simplified in comparison to those applied in EPD. The explanations for this were published in the paper devoted to DPD database (Arkhipova,1995): it was necessary to make the sample more representative in order it would be possibly to classify promoters and to sub-divide them into subsets. As a negative consequence of such way of set extension was some decrease in reliability of TSS positioning. However, it was suggested that some number of entries with false detected TSS will not significantly influence the result of statistical analysis. Incorporation into EnDPD of the data on location of mostly 5'- located ESTs of corresponding genes (this is an independent approach for TSS mapping) favors to decrease of the impact of mentioned disadvantages. It should be noted that currently in analysed set only 63% of promoters have corresponding ESTs in the interval +1 ÷ +100.

Usage of EST for independent verification of promoter localization made possible to refine TSS localization in many cases. If in the sequence of some arbitrary gene, the ending of most 5'-located EST was mapped downstream TSS, known from the literature, then this position was not changed. On the contrary, if the ending of the most 5'-located EST was mapped upstream by the value of ≥ 5 bp, then the novel TSS position was registered in the database. The results of such revision have shown how effective were the criterions formulated by P. Bucher (Бухер,1997): out of 52 EPD originated and supplied with ESTs promoters of *D. melanogaster*, novel TSS positioning was revealed for 11 promoters only. In the sample of *D. melanogaster* promoters, which are not represented in EPD (in total, 156), the revision was made for 33 out of 101 promoters supplied with ESTs. These estimates give evidence how necessary is application of an informational resource such as EST-mapping for detection of TSS position in genes.

One more discriminative feature of EnDPD is its integration into informational systems with the higher level of organization. This is illustrated by the links to the profiles supporting the results of context analysis of promoter and nucleosome potentials, produced by original software program (Levitsky et al., 2000). For examle, in Fig. 2 such profiles for above mentioned second promoter of Rbp9 gene is presented.

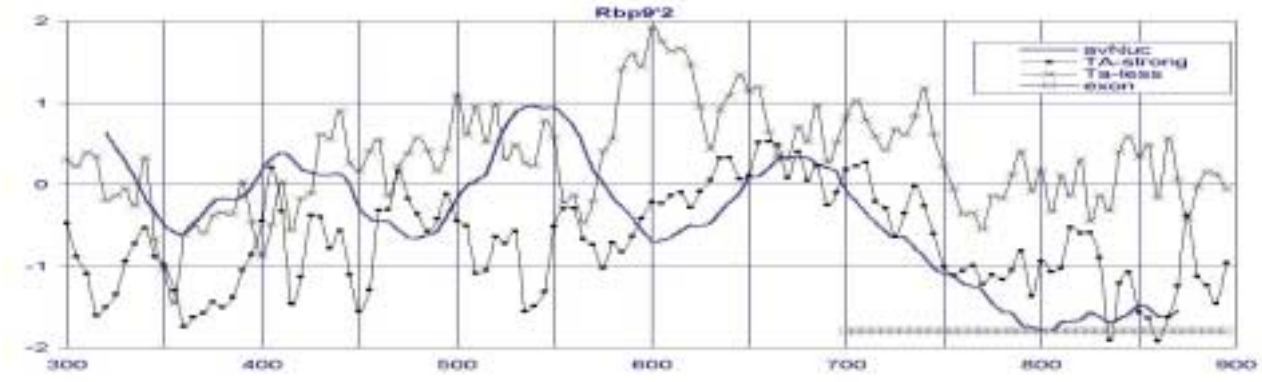


Figure 2. Profiles of recognition functions for nucleosomal site potential and potentials of TATA-strong, and TATA-less promoters calculated for nucleotide sequence of Rbp9 gene second promoter.

In addition, there is a possibility to carry out analogous analysis with customized parameters or analogous analysis of a nucleotide sequences containing novel promoters of *D. melanogaster*. Apparantly, this analysis can be applied to orthologous promoters of other Arthropoda or Invertebratae species in order to explore promoter organization comparatively.

Acknowledgements

The authors are grateful to Galina Orlova for translation of the manuscript into English.

References

1. Arkhipova I.R.(1995) Promoter elements in Drosophila melanogaster revealed by sequence analysis. Genetics, 139: 1359-1369.
2. Bucher P. (1997) Description of eukaryotic promoters in data base EPD. Mol.Biol.(Moscow), 31:616-625.
3. Bucher P. (1999) Regulatory elements and expression profiles. Curr. Opinion in Struct. Biol., 9:400–407.
4. Bucher P. and Trifonov E.N. (1986) Compilation and analysis of eukaryotic POL II promoter sequences. Nucleic Acids Res. 14: 10009-10026.
5. Burke T.W., Kadonaga J.T. (1996) Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev., 10: 11-24.
6. Duret L. and Bucher P. (1997) Searching for regulatory elements in human noncoding sequences. Curr. Opinion in Struct. Biol., 7:399–406.
7. Fickett J.W., Hatzigeorgiou A.G., Eukaryotic promoter recognition. Genome Res., 1997, 7(9), 861-878.
8. Krogh A. (2000) Using database matches with HMMGene for automated gene detection in Drosophila. Genome Res. 10: 483–501.
9. LevitskyV.G., Katokhin A.V. and Kolchanov N.A. (2000) Inherent modular promoter structure and its application for recognition tools development Computational Technologies, 5, Special Issue: 41-47
10. Ohler, U. and Reese, M. R. (1998), Detection of eukaryotic promoter regions using polygrams editor, Molekulare Bioinformatik, Aachen, Shaker (R.Hofestädt (Ed.)): 89-100.
11. Pedersen A.G., Baldi P., Chauvin Y., Brunak S., The biology of eukaryotic promoter prediction--a review. *Comput. Chem.*, 1999, 23, 191-207.
12. Perier R.C., Junier T. and Bucher P. (1998) The eukaryotic promoter database EPD. Nucleic Acids Res.,26: 353-357.
13. Perier R.C., Junier T., Bonnard C. and Bucher P. (1999) The eukaryotic promoter database (EPD): recent developments. Nucleic Acids Res.,27: 307-309.
14. Reese M.G., Kulp D., Tammana H. and Haussler D. (2000a) Genie—Gene finding in Drosophila melanogaster. Genome Res. 10: 529–538.
15. Reese M.G., Hartzell G., Harris N.L., Ohler U., Abril J.F.,and Lewis S.E. (2000b) Genome annotation assessment in Drosophila melanogaster. Genome Res. 10: 523–528.
16. Salamov A.A. and SolovyevV.V. (2000) Ab initio gene finding in Drosophila genomic DNA. Genome Res. 10: 516–522.

EVOLUTIONARY DYNAMICS OF LOW COMPLEXITY DNA SEQUENCES IN CODING AND NON CODING REGIONS

Claudia Acquisti, Giuseppe Mersi, Marcello Buiatti, Patrizia Bogani

Department of Animal Biology and Genetics, Via Romana 19 50125 Firenze,Italy
*Corresponding author

Keywords: phylogenetic tree, phylogenetic reconstruction, low complexity DNA sequences

Resume

Constraints to randomness in the distribution of nucleotides along DNA sequences have widely been studied in order to understand the complex shaping of genomes during evolution.

Since origin of life coding sequences have been selected according to the functional value of the coded proteins, while the fixation of non random sequences in non coding regions is a consequence of other, still partially unknown, features. In this sense deviations from randomness in DNA sequences can be considered as records of the evolutionary history of organisms.

In this work an algorithm have been developed to measure the low complexity (homopurine, homopyrimidine, homo AT, homo GC, 1-2-3 nucleotide repeats) relative weight in genomes of organisms differently located in the phylogenetic tree of life.

It has been shown that the weight of low complexity sequences in genomes greatly increases with evolution. Furthermore non coding and coding regions exhibit different evolutionary behaviours; low complexity weight being higher and increasing much faster in non coding than in coding sequences.

Low complexity analysis of prokaryotic, eucaryotic and organellar genomes, led us a phylogenetic reconstruction generally coherent with known molecular phylogenesis which will be thoroughly discussed in view of a possible functional role of the studied sequence classes also in relation to experimental data obtained in our laboratory on the hypervariable nature of specific low-complexity regions.

THEORETICAL DESIGN OF EFFECTIVE ANTISENSE OLIGONUCLEOTIDES

Sczakiel G.

Institut für Molekulare Medizin, Medizinische Universität zu Lübeck, Germany
e-mail: sczakiel@imm.mu-luebeck.de

Keywords: antisense, oligonucleotides, RNA, RNA-folding, RNA-DNA annealing

Resume

Motivation:

Among the large number of possible antisense species against a given target RNA, only a small number shows effective suppression of the target in living cells. Experimental approaches to identify promising local target sequences and complementary antisense strands, respectively, have provided tools to define antisense oligonucleotides at a higher than statistical probability of success. However, most of the experimental methods are expensive and have basal limitations including the lack of considering adequatly the complete relevant sequence- and structure space. Thus, complementary to their use, computer-based theorectical approaches with their ability to look at complete sets of relevant data seem to provide important information for the design of effective antisense species (Sczakiel, 2000). Here, a concept is described which makes use of structural analyses of computer-predicted target RNA structures to identify promising local targets and, hence, to design potent antisense species. Results from inhibition studies support the usefulness of this approach.

Results:

Here we used a phylogenetic analysis of large sets of secondary strcuture predictions of target RNA in order to select promising local target motives and to defince antisense sequences (Patzel et al., 1999). This theoretical approach led to the design of antisense oligonucleotides with significantly improved efficacy in mammalian cells. It shows the potential usefulness of computerized analyses of computer-predicted RNA structures and its impact on the understanding of the structure-function relationship of antisense oligonucleotides and their targets.

Availability:

The RNA folding algorithms used here (mfold, fold, etc.) are academically and commercially available from many groups. However, it should be noted that the specific folding algorithm is not crucially important for the results of the approach used here.

References:

1. Sczakiel,G. (2000) Theoretical and experimental approaches to design effective antisense oligonucleotides. *Frontiers in Biosciences*, **5**, 194-201.
2. Patzel,V., Steidl,U., Kronenwett,R., Haas,R. & Sczakiel,G. (1999) A theoretical approach to select effective antisense oligodeoxyribonucleotides at high statistical probability, *Nucleic Acids Res.*, **27**, 4328-4334.

PERIODICITIES IN THE PRIMARY STRUCTURE OF *E.COLI* PROMOTERS

Kravatskaya G.I., Frank G.K., Makeev V.Yu., Polozov R.V., Esipova N.G.

Institute of Molecular Biology by V.A. Engelgardt of RAS, Moscow, Russia
*Corresponding authhor
e-mail:gk@imb.ac.ru, nge@imb.ac.ru, polozov@venus.iteb.serpukhov.su
gleb@imb.ac.ru, makeev@imb.ac.ru

Keywords: transcription, RNAP, promoters, classification, matrix Fourier-analysis

Resume

Motivation:

We believe that an ability of some DNA sequence to be recognized by the RNA-polymerase as promoter is based not only on the "local" features (the so-called "-35", "-16", and "-10" regions), but also on the more global characteristics of DNA structure (i.e., heterogeneity of hydrated DNA covering, bending, superspiralization, etc.), which are related to periodic localization of nucleotides and their groups. In order to study periodicity in the promoter structure and their role in the process of transcription initiation, we have applied the matrix Fourier analysis.

Results:

Promoter nucleotide sequences contain periodicities. The same observation is true for localization of the contacts with RNA-polymerase within the promoters. The variability of Fourier spectra of promoter sequences is less than variability of their primary structures. There exist the groups of promoters with analogous Fourier-spectra. We have found that nucleotide sequences of all promoters contain the periodicities that are not coincident to canonical periodicity (10.5 bp) of the sugar-phosphate carcass of the B-DNA, which destables the DNA duplex and makes an impact into formation of the open complex.

Introduction

It is unknown up to now, how is the specificity of recognition of promoters by RNA-polymerase produced. In the work given, we suggest an alternative methodology for analysis of promoter sequences and a novel approach to analysis of localization of the contacts between RNA-polymerase and promoters.

Methods and algorithms

The matrix Fourier analysis [Makeev et al., 1996] was used for revealing periodicities in the primary structure of promoters and in localization within promoter the points of interaction with the RNA-polymerase. These points of contacts were experimentally detected elsewhere [Chenchick et al., 1981, Duval-Valentin et al., 1986]. For promoter classification on the basis of their Fourier-spectra, we have applied the cluster-analysis [Statistica" V.5.0] .

Implementation and results

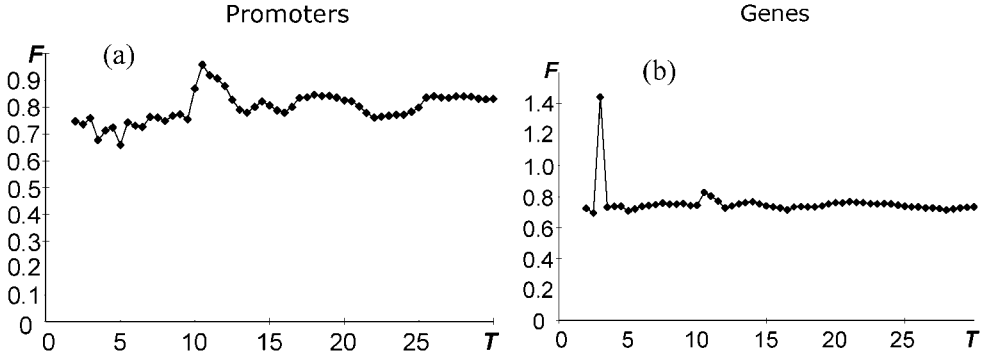


Fig. 1. Average alteration of the spectral power (*F*) in dependence upon the period (*T*) for the sets of promoters (a) and the fragments of the coding regions (b) in *E.coli*

By means of the program PERF, we have obtained the Fourier-spectra of promoter sequences according to location of different nucleotides and dinucleotides. Besides, a comparison with Fourier-spectra from the other genome regions was made (Fig.1). Only for 56% of promoters from [Lisser& Margalit, 1993], we have detected prnounced peaks corresponding to the period T=10,5. Within the spectra of 57% promoters, the peaks are clearly seen that correspond to the period T=11 nucleotides. In nucleotide sequences of all promoters, we have found periodicities that are not multiple to canonical periodicity (10,5 bp) of the sugar-phosphate carcass of B-DNA. In addition, for several promoters, we have got the Fourier-spectra of distribution of the points of interaction between RNA-polymerase and promoter.

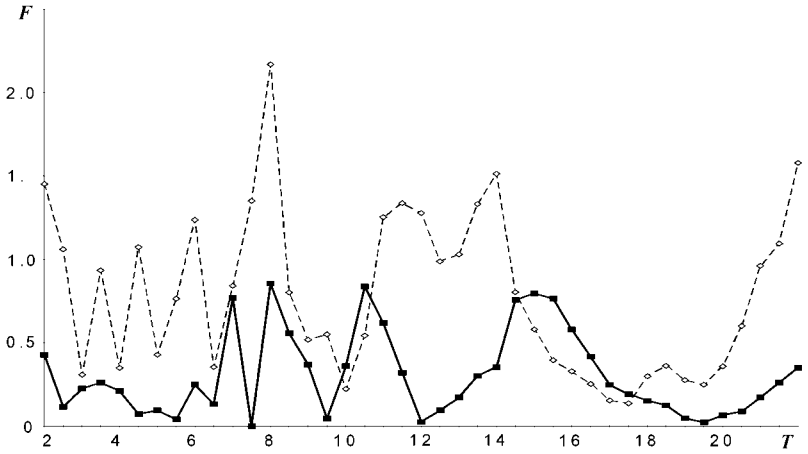


Fig. 2. Alteration of the spectral power (*F*) in dependence of the period (*T*), due to Fourier transformation or the lac UV5 promoter in *E.coli* according to distribution of all four nucleotides (continuous line) and the points of contacts of complete RNA-polymerase (dotted line)

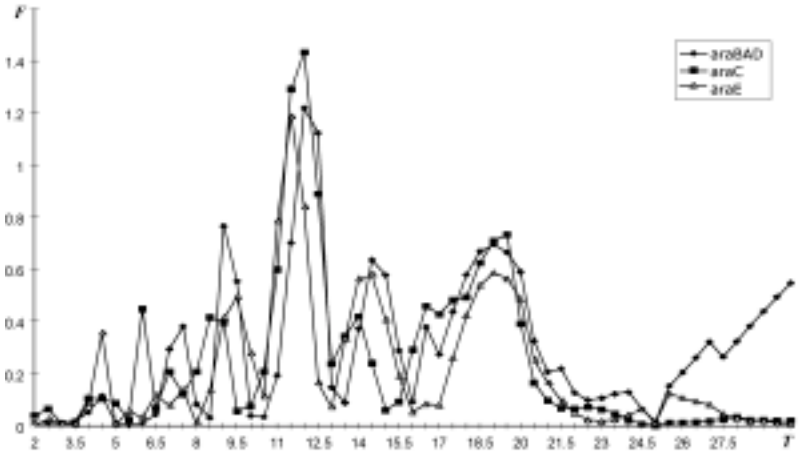


Fig. 4. Fourrier-spectra for the araC, araE, araBAD promoters according to distribution of the dinucleotides AA/TT.

This work is supported by the RFBR grant N 00-04-48351.

Discussion

We suppose that periodicities that are not multiple to the period of the BDNA may destabilize the double DNA helix and make an impact into formation of the open complex of the RNA-polymerase with promoter. The Fourier-spectra according to the localization of the points contacting with RNA-polymerase within several promoters have some similar features, except the discrepancies. We have noted the presence of periodicity by 7 nucleotides (and multiple to this periodicity) in the averaged for several promoters spectrum of localization of these contacts. Nucleotide sequences of 43% of promoters have clear periodicity equalling to 7 nucleotides. The variability of Fourrier-spectra of promoter sequences is less than the variability of their primary structures. We have observed the groups of promoters with the similar Fourrier-spectra. It should be noted that several groups of promoters with the similar Fourrier-spectra are regulated by one and the same proteins [eg., araC, araE, and araBAD promoters are regulated by the CRP and araC proteins (Fig. 4)]. The similarity of the sequences of these promoters equals to 30%, although the Fourrier-spectra of these promoters (according to localizations of the AA/TT dinucleotides) are very similar. Nucleotide sequences of promoters contain periodicities. The same observation is true for localization of the contacts of the RNA-polymerase with promoters. Possibly, RNA-polymerase looks for the specific periodic "motifs", which may be different for different groups of promoters.

References

3. Makeev, V.Ju., Tumanyan, V.G. (1996) Search of periodicities in primary structure of biopolymers: a general Fourier approach. *CABIOS* 12, 49-54
4. Chenchick, A., Beabealashvilli, R., Mirzabekov, A. (1981). Topography of interaction of Escherichia coli RNA polymerase subunits with lac UV5 promoter. *FEBS Lett*. 128, 46-50
5. Duval-Valentin, G., Ehrlich, R. (1986). Interaction between *E.coli* RNA polymerase and tetR promoter from pSC101: homologies and differences with other *E.coli* promoter systems from close contact point studies. *Nucleic. Acids Res.* 14, 1967-1983

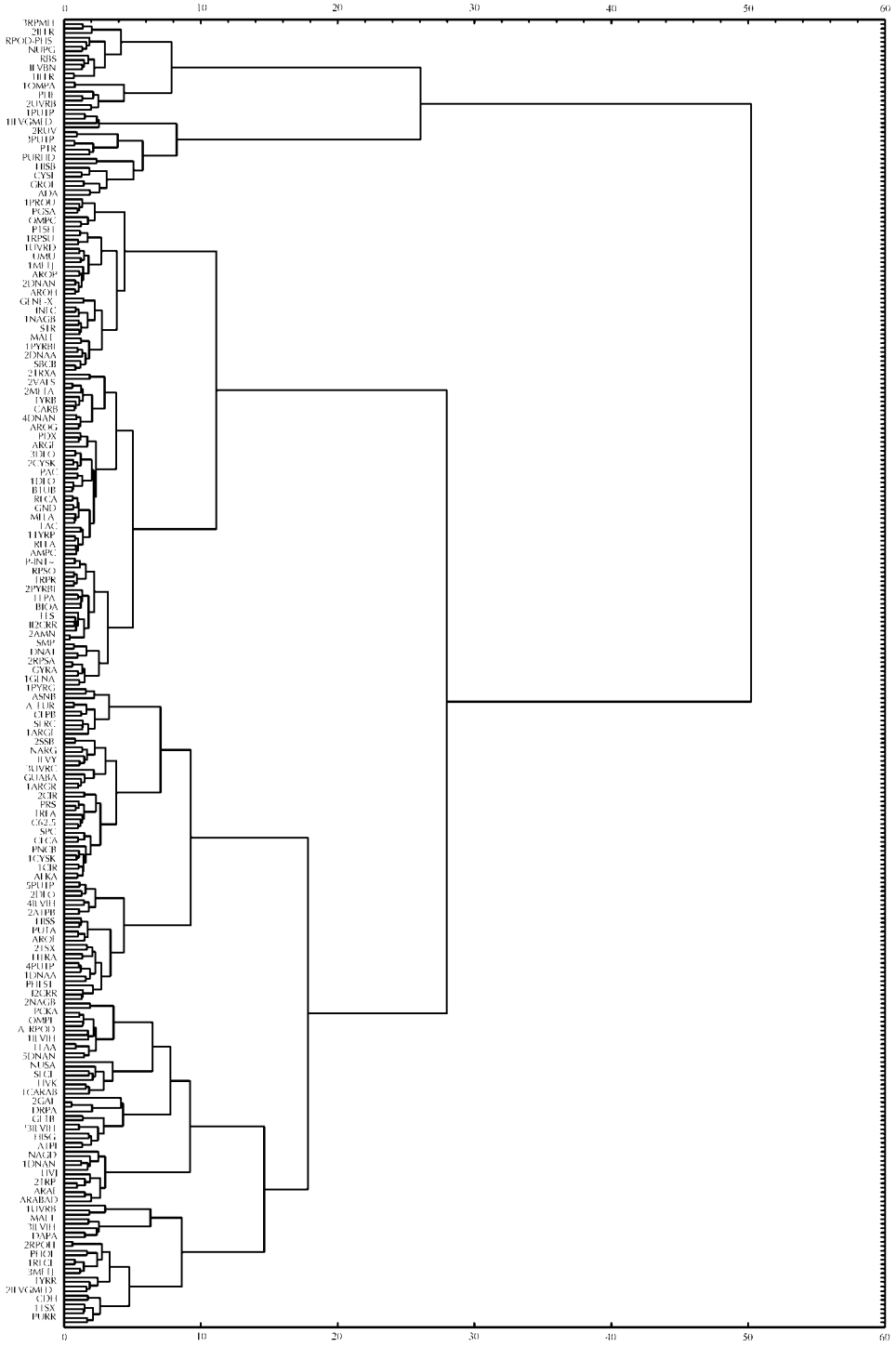


Fig 3. The result of the cluster analysis of 290 promoters of *E.coli* from the compilation [Lisser, Margalit, 1993] on the basis of their Fourier spectra (for AA/TT dinucleotides)

MOLECULAR MODELLING OF DISEASE-CAUSING SINGLE-NUCLEOTIDE POLYMORPHISMS IN COLLAGEN

Milchevsky Ju.V., Ramensky V.E., Zorov B.S., Esipova N.G., Tumanyan V.G.

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov str., Moscow, 117984 Russia

Abstract

The purpose of the work was to investigate at the molecular structural and energy levels the consequence of amino acid substitutions in collagen that cause systemic diseases. The data have been systematized on defects in human collagen III, and the patterns of single-nucleotide polymorphisms collected; then molecular mechanics calculations were performed for native and mutant collagen molecule fragments. The observed energy components and structural alterations that accompany particular amino acid substitutions were used to propose an interpretation of negative consequences in terms of stability and hydration of the macromolecule.

This work was supported by the Russian Foundation for Basic Research (grant 98-04-48755).

ELUCIDATION OF THE STRUCTURE OF WHOLE PROTEIN: HUMAN COLLAGEN III

Milchevsky Ju.V., Zorov B.S., Esipova N.G., Tumanyan V.G.

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov str., Moscow, 117984 Russia

Abstract

The aim of the work was to elucidate the molecular structural and energy characteristics of three-stranded collagen macromolecule. The computations have been performed on the basis of primary structure of human collagen III by improved method of molecular mechanics. The previously elaborated model with two nets of hydrogen bonds for tripeptides with the amino acid in the second position of the tripeptides was used. The common one bonded model was used for tripeptides with imino acid in the second position of the tripeptide. By optimization energy taking into account flexibility of pyrrolidine cycles realistic structure of the protein have been attained. The role of various components of energy including energy of hydration and ionic ionic bonds along the fibrous molecule is discussing. Thus, it is deduced for the concrete protein how a one-dimensional structure determines three-dimensional structure.

This work was supported by the Russian Foundation for Basic Research (grant 98-04-48755).

AUTHOR INDEX

S

T

V

Z

KEYWORDS INDEX

S

T

V

W

X