STATEMENT OF RESEARCH AND TEACHING INTERESTS (PAVEL BARANOV)

RESEARCH BACKGROUND

Graduate studies.

I have completed my Ph.D. studies in Moscow State University (Moscow, Russia) in the laboratory of Profs Alexei Bogdanov and Olga Dontsova in collaboration with a group of Dr. Richard Brimacombe that involved visits to Max Planck Institute for Molecular Genetics (Berlin, Germany). The main interest of this laboratory was cross-linking studies of the ribosome structure. My Ph.D. project was to develop a method that would allow insertion of a photolabel into a desired position of ribosomal RNA for the following cross-link induction and its analysis. I developed such a method based on site-directed cleavage of RNA by RNAse H in the presence of 2'-OMe chimeric oligonucleotides and subsequent ligation of photolabeled nucleotide into the site of the cleavage.

During this time, I initiated a bioinformatics project, the Database of Ribosomal Cross-links (DRC) available at http://www.mpimg-berlin-dahlem.mpg.de/~ag_ribo/ag_brimacombe/drc/defaultlow.html. The purpose of this database was to provide researchers with a compilation of categorized information about cross-linking studies of the ribosome. This database is currently not updated since the crystal structures of bacterial ribosome subunits have been solved. However, it is still popular among researchers that use cross-linking studies for other ribonucleoproteides, because comparison of the data present in DRC with the real 3D structure of the ribosome provides a valuable information on reliability of different photo reagents and crosslinking techniques.

I also played an active role in computer modeling of 23S ribosomal RNA topology inside 50S ribosomal subunit using cryo-EM and biochemical data. I have developed a chime-based web resource for the visualization of this model. It is available at http://www.mpimg-berlin-dahlem.mpg.de/~ag_ribo/ag_brimacombe/drc/ribosome/.

Therefore, during my graduate studies I obtained a solid background in the fields of Biochemistry, Structural Biology and Bioinformatics.

Postdoctoral studies.

After completing my Ph.D and a very short postdoctoral training at the group of Dr. Richard Brimacombe, I joined the laboratory of Drs. John Atkins and Raymond Gesteland. The main interest of this laboratory is 'recoding', a term used to describe utilization of non-standard translation (such as ribosomal frameshifting, stop codon readthrough, selenocysteine insertion, translational bypass, etc) for gene expression purposes.

I used my knowledge of ribosome structure and translational mechanisms to understand how non-conventional translation occurs. I have developed a new reductionistic model of ribosomal frameshifting, which provides a unified molecular mechanism for different types of frameshifting and is supported by experimental data.

My time in the Atkins/Gesteland lab coincided with explosive development of the genomic field. I was fascinated by opportunities that were produced for comparative sequence analysis and I decided to utilize these opportunities to study different aspects of recoding. I have identified a number of new bacterial and eukaryotic genes that utilize ribosomal frameshifting for their expression. I also used comparative sequence analysis in combination with site-directed mutagenesis to identify previously unknown stimulatory signals embedded in mRNA to stimulate recoding.

Therefore, during my postdoctoral studies I expanded my expertise in Genetics, Bioinformatics and Genomics.

Currently, I direct several research projects. This includes

- identification of new recoding events in *D. radiodurans* using bioinformatics predictions and proteomics mass-spectrometry data in collaboration with Dr. Richard Smith laboratory (Pacific Northwest National Lab);
- elucidation of the nature of certain ('mutation'/'sequence_errors'/'potential recoding events') in *Shewanella oneidensis* in collaboration with Dr. Margaret Romine (Pacific Northwest National Lab);
- investigation of several basic aspects of viral frameshifting using coronaviruses in collaboration with Dr. Michael Howard (University of Utah).
- I also coordinate the RECODE database, the web resource that was created and is currently maintained by several research groups from different countries. RECODE is available at http://recode.genetics.utah.edu.

FUTURE PLANS

As an independent investigator, within next five years, I plan to establish an active extramurally funded research program that will focus on two overlapping issues related to recoding. 1. Detection of new recoding cases (short terms) 2. Elucidation of recoding mechanisms and application of its knowledge to modulate non-standard decoding (long terms). Contemporary bioinformatics and experimental approaches will be used to address these questions.

1. Detection of new recoding cases (short terms)

Introduction. The RECODE database mentioned above lists about 350 genes that are known to utilize recoding (this number is likely to double after year 2004 update). However, there are less than 50 different types of genes (phylogenetically and functionally unrelated). The majority of these genes were found during studies of the expression of individual genes, so that finding of recoding cases was almost coincidental. Recent enormous progress in genome sequencing and annotation does not directly contribute to the growth of our knowledge about new recoding events, because there are no robust computational algorithms to predict such events. Moreover, many annotated genomes still omit proper annotations of genes that are very well known to require recoding for their expression. Within the last three years, a few very successful attempts were performed to find new cases of recoding using systematic approaches. The results of these studies indicate that indeed this is a largely unexplored area and the currently known cases constitute just a tiny proportion of those yet to be discovered. The discovery

of the 22nd amino acid pyrrolysine two years ago suggests that even our current knowledge of the genetic code may be incomplete.

Brief research design. The following is a brief description of a few approaches that can be applied to search for new recoding cases.

There are a few highly conserved genes known to utilize recoding, such as bacterial release factor 2 or eukaryotic antizymes. A remarkable conservation of frameshift sites in them and a large number of known sequences allows creation of probabilistic models for predicting additional orthologous. Such probabilistic models can be used for automatic detection and annotation of recoding cases in newly sequenced genomes (currently this is performed manually and requires a certain expertise. As a result, a number of cases are mis-annotated). The disadvantage of such an approach is that it is applicable only to those types of genes where recoding was already documented for their orthologous.

A number of short sequences (6-9 nucleotides) are capable of triggering non-conventional decoding in certain organisms. Therefore, searches for such sequences will be performed. Whether a particular occurrence of such a sequence has a functional importance will be assessed with comparative sequence analysis and following experimental verification. Unfortunately, the nature of such sequences differs among organisms because of differences in the translational apparatus (e. g. tRNAs and their relative concentrations are highly variable among different species). As a result, this approach can be used only for organisms with prior knowledge of their non-conventional decoding rules.

In order to apply similar method to other organisms, it is possible to perform statistical analysis of their genomes. Sequences that trigger recoding are harmful if they occur without a purpose (e. g. frameshift-prone sequence in a gene decoded in the standard manner will result in a synthesis of some aberrant protein product). As a result such sequences are underrepresented in coding regions of genome (this is particularly evident for highly expressed genes). This feature can be used to derive putative recoding sequences whose “recoding nature” can be assessed using experimental approaches.

Although short shift-prone sequences are capable of triggering unconventional decoding at low efficiency, much higher efficiency is usually required for gene expression purposes. Therefore, special sequence elements often evolved in mRNA to modulate efficiency of recoding. These *cis*-acting elements are known as recoding stimulators. Recoding stimulators can function at different levels. The majority are known to function as complex RNA structures that interact with components of the translation machinery to affect the decoding process. In addition, there are stimulators that interact with ribosomal RNA *via* complementary interactions and stimulators encoding nascent peptides capable of affecting ribosome structure. Also, there are a number of *cis*-elements whose nature remains unclear. Nucleotide sequences encoding such stimulators are usually conserved and therefore detectable *via* comparative sequence analysis. Consequently, it is possible to search for recoding stimulators in the genomic sequences and verify whether they are used for recoding purposes.

Several other approaches can be applied for this purpose and include applications of proteomics data and genetic manipulations with translation machinery of a given organism.

2. Mechanisms of non-standard decoding and modulation of non-standard decoding (long terms)

Certain mRNAs manipulate genetic decoding in a remarkably audacious manner, they program ribosomes to shift reading frames during translation, redefine codon meaning or skip an mRNA segment without translation (a unique example of ribosomal bypassing is in decoding phage T4 gene *60*). Such natural ability provokes the thought that we can learn how to change gene decoding and how to control it, if we understood how recoding is achieved in nature. There are at least two most obvious applications of such ability.

1. Recoding is frequently used for the expression of viral genes and virulence genes in bacterial pathogens. A proof of the importance of this mechanism is the astonishing conservation of sequences responsible for recoding. For example, despite its known variability, HIV has invariant frameshift site U_UUU_UUA. Therefore, frameshifting in HIV and other viruses can be used as a therapeutic target for development of new and powerful antiviral drugs.

2. A number of genetic diseases are caused by single nucleotide mutations, such as frameshift and nonsense mutations. Certain compounds, such as aminoglycoside antibiotics are known to lower translation accuracy in mammals and can be potentially used to elevate unconventional translation of mutant genes. Currently NIH supports several projects that proposed application of aminoglycoside antibiotics to treat severe genetic diseases such as cystic fibrosis, muscular dystrophy and hemophilia. The major problem of these approaches is toxicity. Aminoglycosides target translational machinery (not only) and, as a result, they affect gene readout in general causing some mistranslation of all mRNA in the cell. This problem can be avoided if we learn how to use a specific mRNA of interest as a target.

Several questions should be answered about the mechanism of non-standard decoding before we will be able to manipulate it. What is happening during +1 frameshifting inside the ribosome? Does a tRNA income into A-site in the +1 frame directly or +1 frame is set by P-site tRNA slippage in the ribosome? Do A-site and P-site tRNAs slip simultaneously during -1 frameshifting or do they slip sequentially?

A number of recoding stimulators act at the level of RNA structures, such as strong stem-loops and RNA pseudoknots. Several theories (not necessarily mutually exclusive) have been proposed to explain their effect and they all assume that such structures should be formed at least for some period of time to stimulate recoding in the ribosome. However, it is unclear in what proportion of mRNA these structures are actually formed. How is formation of these structures affected by the density of ribosomes translating mRNA? Why are not all of these structures interchangeable? This is a tiny proportion of all questions that merit investigation.

In conclusion, I want to point out that my exact future research plans are highly dependable on my new research environment and the opportunities to collaborate with other researchers.

DESCRIPTION OF TEACHING EXPERIENCE

While being a graduate student, I have been responsible for the organization of facultative course in Molecular Biology for undergraduate students in School of Chemistry, Moscow State University, Russia.

I supervised a practical laboratory course “Application of radioactive labels in Molecular Biology” for Diploma students at the Department of Chemistry and Biochemistry of Natural compounds.

I co-supervised two Diploma students, Olga Gurvich and Andrei Leonov. Olga Gurvich is currently a graduate student at the Molecular Biology Program, University of Utah. Andrei Leonov is recently obtained his Ph.D. degree at Moscow State University.

As a postdoctoral researcher at Human Genetics Department, University of Utah, I continue to supervise research activities of several undergraduate students and rotation graduate students.