T A B I S 2 0 1 0

Theoretical Approaches to Bio-Information Systems

Institute of Physics, Belgrade

BOOK OF ABSTRACTS

Belgrade 20-21. 05. 2010.

Organizing Committee

- Branko Dragovich (Chairman),
  Institute of Physics, Belgrade
- Gordana Pavlović-Lažetić,
  Faculty of Mathematics, Belgrade
- Zoran Rakić,
  Faculty of Mathematics, Belgrade
- Paul Sorba,
  Lapth/CNRS, Annecy, France

Electronic Structure of Nucleotides Interacting with Nanotube Leads

V. Cerovski and R. Žikić

Institute of Physics Belgrade, Pregrevica 118, 11080 Belgrade, SERBIA

Abstract.

Electronic structure of A,C, G i T nucleotides interacting with 3-3 Carbon nanotube (CNT) leads is numerically studied using the Density Functional Theory in the context of recently proposed experimental method for fast DNA sequencing. Electronic structure of CNT and A,C,G i T are presented first, followed by a simple analysis of the integrated density of states near the Fermi energy of nucleotide+CNT leads systems, giving qualitative estimates of the electronic transport properties at finite bias of each of the four nucleotides studied.

p-Adic Modeling of the Genetic Code

Branko Dragovich
Institute of Physics, Pregrevica 118,
11080 Zemun, Belgrade, SERBIA

Abstract.

The genetic code is a mapping of 64 codons onto 20 amino acids and one stop signal. In p-adic approach, codons are presented as natural numbers expended in the base 5 with three digits related to nucleotides: C=1, A=2, U=T=3, G=4. With respect to the smallest 5-adic distance between codons, they clasterize into 16 quadruplets, which under 2-adic distance decay into two doublets. In the vertebral mitochondrial code, two codons inside these doublets determine the same amino acid or stop signal. Other genetic codes can be regarded as small deformations of the this mitochondrial one. Some other aspects of the p-adic genetic code will be also presented.

The Logic Of Scientific Discovery
- Case Study Of DNA -

Petar Grujić
Institute of Physics, POB 57, 11080 Belgrade, SERBIA

Abstract.

Great discoveries are usualy product of an interplay of chance and necessity. The case of establishing the primary and secondary structures of deoxiribonucleic acid is not an exception. Many laboratories with numerous people have conributed to decoding DNA structure, in a direct or indirect way. We review briefly the history of this discovery and discuss a number of epistemological issues, concerning an interplay of the theory, models and empirical evidence. In particular the role of the theory of X-ray crystallography is emphasized, and the epistemological comparison between bottom-up and top-down methodological approaches is made too.

2-Adic parametrization of the genetic code

S.V. Kozyrev

Steklov Mathematical Institute, Moscow, Russia.

Genetic code is the map which puts in correspondence to a codon, i.e. a triple of nucleotides (denoted by A, U, G, C) an amino acid. There are 64 possible codons. In [1] the 2-dimensional 2-adic parameters on the set of codons were introduced. This parametrization arranges the codons in the 8×8 table (the 2-adic plane)

AAA	AAG	GAA	GAG	AGA	AGG	GGA	GGG
AAU	AAC	GAU	GAC	AGU	AGC	GGU	GGC
UAA	UAG	CAA	CAG	UGA	UGG	CGA	CGG
UAU	UAC	CAU	CAC	UGU	UGC	CGU	CGC
AUA	AUG	GUA	GUG	ACA	ACG	GCA	GCG
AUU	AUC	GUU	GUC	ACU	ACC	GCU	GCC
UUA	UUG	CUA	CUG	UCA	UCG	CCA	CCG
UUU	UUC	CUU	CUC	UCU	UCC	CCU	CCC

with the 2-adic norm on this table.

Application of the genetic code to the above 2-adic plane gives the table of amino acids on the 2-adic plane

$\frac{\text{K}}{\text{N}}$	$\frac{\text{E}}{\text{D}}$	$\frac{\text{R}}{\text{S}}$	G
$\frac{\text{Ter}}{\text{Y}}$	$\frac{\text{Q}}{\text{H}}$	$\frac{\text{Ter}\vert\text{W}}{\text{C}}$	R
$\frac{\text{I}\vert\text{M}}{\text{I}}$	V	T	A
$\frac{\text{L}}{\text{F}}$	L	S	P

where Ter is the stop codon and we use the standard 1-letter notations for the amino acids. Each square in the above table is the 2×2 square in the 2-adic plane of codons.

The introduced 2-adic parametrization allows to express the degeneracy of the genetic code as the local constancy of the map of the 2-adic argument. Moreover, the physical–chemical properties of the amino acids (i.e. hydrophobicity) are clustered in the 2-adic plane. We also discuss the relation of the above parametrization to the Rumer symmetry of the genetic code.

References

[1] A.Yu. Khrennikov, S.V. Kozyrev, Genetic code on the dyadic plane. Physica A: Statistical Mechanics and its Applications. 2007. V.381. P.265-272. arXiv:q-bio.QM/0701007

[2] B.Dragovich, A.Dragovich, A p-Adic Model of DNA Sequence and Genetic Code. p-Adic Numbers, Ultrametric analysis and Applications, 2009. V.1. N.1 P.34-41. arXiv:q-bio/0607018v1

[3] Yu.B. Rumer, On systematizing codons in the genetic code. Doklady Akademii Nauk SSSR. 1966. V.167. N.6. P.1393-1394. (In Russian)

Cancer diagnosis based on microarray gene expression data

Damjan Krstajić
Research Centre for Cheminformatics (RCC), Belgrade, SERBIA

Abstract.

In the past decade, new experimental technologies in the field of genomics have resulted in a wealth of biomedical data. The new types of data, i.e. gene expressions, have small number of observations and high number of features. They are referred to as "high-dimensional" by statisticians, and pose a challenge to bioinformaticians. In our talk wewill give a summary of published statistical approaches which use gene expression data to build models for predicting survival or type of cancer.

Reexamination of Correlations of Amino Acids with Particular Secondary Structures

Results by: Saša N. Malkov, Miodrag V. Živković, Miloš V. Beljanski, Srđan D. Stojanović, Snežana D. Zarić, Michael B. Hall

Presenter: Saša N. Malkov

Abstract.

Using the data from Protein Data Bank the correlations of primary and secondary structures of proteins were analyzed. The correlation values of the amino acids and the eight secondary structure types were calculated, where the position of the amino acid and the position in sequence with the particular secondary structure differ at most 25. Clear preferences of amino acids towards certain secondary structures classify amino acids into four groups: α-helix preferrers, strand preferrers, turn and bend preferrers, and *His* and *Cys* (the latter two amino acids do not show clear preference for any secondary structure). Amino acids in the same group have similar structural characte-ristics at their Cβ and Cγ atoms that predicts their preference for a particular secondary structure. All α-helix preferrers have neither polar heteroatoms on Cβ and Cγ atoms nor, branching nor aromatic group on the Cβ atom. All strand preferrers have aromatic groups or branching on the Cβ atom. All turn and bend preferrers have polar heteroatom on Cβ or Cγ atoms or do not have a Cβ atom at all. These new rules can be helpful in making predictions about non-natural amino acids. The correlations for different distances are significant at distances between -9 and 10. The results show that the substituents on Cβ or Cγ atoms of amino acid play major role in their preference for particular secondary structure at the same position in the sequence, while the polarity of amino acid has significant influence on α-helices and strands at some distance in the sequence. The diagrams corresponding to polar amino acids are noticeably asymmetric.

n-Gram prediction of genomic islands

N. S. Mitić[1], G. M. Pavlović-Lažetić[1], and M. V. Beljanski[2]

[1] Faculty of Mathematics, Department of Informatics, University of Belgrade, Studentski trg 16, Belgrade, Serbia
nenad|gordana@matf.bg.ac.rs
[2] Institute of General and Physical Chemistry, Studentski trg 12-14, Belgrade, Serbia
mbel@matf.bg.ac.rs

Abstract. Identification of bacterial genome segments known as genomic islands (GIs) in bacterial genomes is an important task since many of them represent inserts that may contribute to bacterial evolution and pathogenesis. Two novel n-gram-based method for prediction of GI will be presented. The first one is based on n-gram frequency distribution. N-gram distributions were combined by union and intersection and two measures defined for evaluating the results - recall and precision. Using the best criteria (by training on the Escherichia coli O157:H7 EDL933 genome), GIs were predicted for 14 Enterobacteriaceae family members and for 21 randomly selected bacterial genomes. These predictions were compared with results obtained from two existing databases of genomic and pathogenecity islands. The results obtained show that application of n-grams improves both relative precision and recall. The results obtained by this method is published in article [1].

The second method is based on using Data Mining classification techniques for predicting GIs. Method is still under development. Various algorithms were tested and generated results are at least good as those obtained by other (non Data Mining) methods.

References

1. N.S. Miticć, G.M. Pavlovicć-Laeticć, M.V. Beljanski: Could n-gram analysis contribute to genomic island determination?, Journal of Biomedical Informatics, 41(6) (2008) 936-943.
2. G.M. Pavlovicć-Lažeticć, N.S. Miticć, M.V. Beljanski: n-Gram characterization of genomic islands in bacterial genomes, Computer methods and programs in biomedicine, 93 (2009) 241-256.
3. N-gram classification and prediction genomic islands, Internal communication

Bioinformatics analysis of genome sequences

Gordana Pavlović-Lažetić

Faculty of Mathematics, University of Belgrade
email: gordana@matf.bg.ac.rs

Abstract. Results will be presented in application of informatics methods in analysis and prediction of genome sequence composition and structure. Reported research has been performed within the bioinformsatics group of the project "Automated Reasoning and Data and Text Mining" (MNTR 144030) and the previous project "Bioinformatics" (MNTR 1856), at the Faculty of Mathematics, financially supported by the Ministry of science and environmental protection of the Republic of Serbia.
First, comparative analysis of the SARS-CoV genome isolates will be presented aimed at establishing genome polymorphism and variant evolution. Then an n-gram-based method will be presented for analysis and prediction of bacterial genome segments known as genome islands (GI). GIs are important because many of them represent insertions contributing to bacterial evolution and pathogenesis.
Some ongoing research will be shortly commented at the end.

References

1. G.M. Pavlović-Lažetić, N.S. Mitić, M.V. Beljanski: n-Gram characterization of genomic islands in bacterial genomes, Computer methods and programs in biomedicine, 93 (2009) 241-256
2. N.S. Mitić, G.M. Pavlović-Lažetić, M.V. Beljanski: Could n-gram analysis contribute to genomic island determination?, Journal of Biomedical Informatics, 41(6) (2008) 936–943
3. G.M. Pavlović-Lažetić, N.S. Mitić, M.V. Beljanski: Bioinformatics analysis of SARS coronavirus genome polymorphism, BMC Bioinformatics. 2004; 5: 65
4. G.M. Pavlović-Lažetić, N.S. Mitić, A.M. Tomović, M.D. Pavlović, M.V. Beljanski: SARS-CoV Genome Polymorphism: A Bioinformatics Study, Geno. Prot. Bioinfo. Vol. 3 No. 1 February 2005, 18-35

A Systemic-Chemical Approach To The Genetic Code

Miloje M. Rakočević

Abstract.

This communication presents several key results from my researches of the genetic code. All of these results support the hypothesis on a complete genetic code, which I expressed explicitly, in an article in the 2004th year [J. Theoret. Biol. 229 (2004) 221-234]; complete code, in the sense that the genetic code was in pre-biotic times and areas as it is today, consisting from four amino bases and 20 amino acids.

A symmetry approach for modelling the genetic code

Paul Sorba
LAPTH/CNRS (France)

Abstract.

The quantum algebra Uq(Su(2)+Su(2)in the limit q=0 is proposed as a symmetry algebra for the genetic code. The nucleotids triplets (codons) in the DNA chain are classified in crystal bases and an operator ensuring the correspondence between codons and amino-acids is constructed from the above algebra. As a first set of applications, correlations for codon usage for quartets and sextets are determined, fitting naturally in the framework of this model, and sum rules are explicited. Then a set of relations between the physico-chemical properties of the amino-acids are derived and compared with the experimental data.