

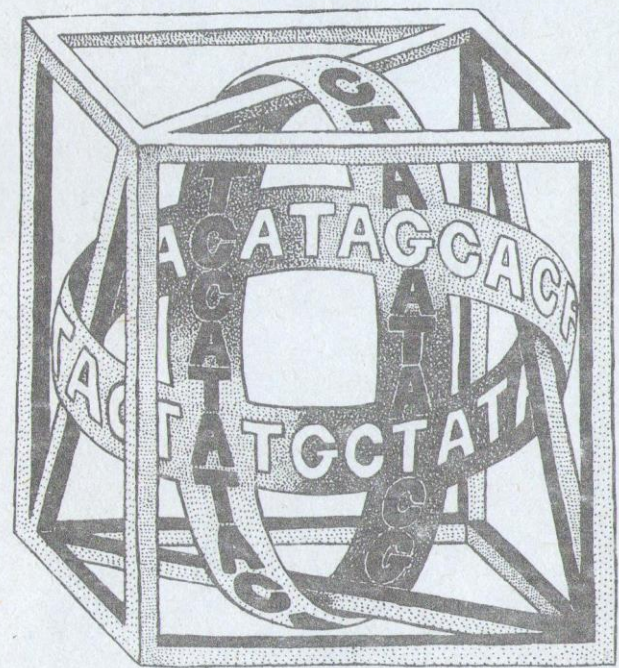
575
M78

ТРУДЫ
Сотруженников ИЦиГ

THE USSR ACADEMY OF SCIENCES
SIBERIAN BRANCH
INSTITUTE
OF CYTOLOGY AND GENETICS

MODELLING AND COMPUTER METHODS
IN MOLECULAR BIOLOGY AND
GENETICS

ABSTRACTS OF THE INTERNATIONAL CONFERENCE



NOVOSIBIRSK. 1990

THE USSR ACADEMY OF SCIENCES
SIBERIAN BRANCH

INSTITUTE
OF CYTOLOGY AND GENETICS

MODELLING AND COMPUTER METHODS
IN MOLECULAR BIOLOGY AND
GENETICS

ABSTRACTS OF THE INTERNATIONAL CONFERENCE

EDITED BY N.A. KOLCHANOV AND V.A. RATNER

NOVOSIBIRSK. 1990

ЭМ
Самодельный

ТРУДЫ
СОТРУДНИКОВ ИЦИГ

57B.1:577.21.004



8/4

INTRODUCTION

The present book contains reports of the International Conference "Modelling and Computer Methods in Molecular Biology and Genetics" which took place from August 27 to September 1, 1990 in Novosibirsk (USSR) on the initiative of the Institute of Cytology and Genetics, Siberian Branch of Academy of Sciences of the USSR.

The Conference was held under the organizational support of Siberian Branch of Academy of Sciences of the USSR, General Biology Department of the USSR Academy of Sciences, the Complex Program of Scientific and Technical Progress (The 5.5.5.3 Project) with the countries members of the Council of Mutual Economic Assistance, All-Union scientific-technical program "Geninform" and the National "Human Genome" Project.

The materials presented correspond to the following fields:

- Computer analysis of nucleotide sequences;
- Computer analysis and modelling of proteins structure;
- Molecular Evolution Theory;
- Molecular-genetic systems' modelling;
- Molecular-genetical and molecular-biological data banks;
- Program packages for molecular-genetical and molecular-biological data analysis;
- New computer-informational techniques in molecular biology and genetics, including expert systems, artificial intelligence systems and knowledge bases.

This book gives a profound outlook on the elaborations and scientific investigations in the mentioned above fields of bioinformatics in 1990.

We hope it will be interesting for specialists, engaged in bioinformatics; mathematical simulation and theoretical examining of macromolecules' (DNA, RNA and proteins) structure, functioning and evolution.

The book can also attract attention of the scientists, interested in applying mathematical and computer analysis to molecular biology, molecular genetics and biophysics.

Dr. N.A. Kolchanov
Prof. V.A. Ratner

L. Milanese , G. De Bellis , R. Andreoli

Institute of Biomedical Advanced Technology of C.N.R., Via
Ampere 56, 20131 Milano ITALY

Using the fluorescent automated sequencer is particularly important to analyze the final sequence obtained from the instrument because of the presence of ambiguity and errors related to the raw data software processing . We have developed a computer program written in Fortran 77 using GKS standard graphic library, running on VaxStations based on VMS operative system. This program allows an easy handling of the sequences and related raw data produced by a fluorescent automated sequencer. It is based on menus and mouse driven choice. It allows to review the sequences linked with the raw data , in order of making the final corrections. It is also possible to compare the results obtained from two different runs of the same sequence, in presence of both the sequences and the related raw data, in order to obtain the best final sequence. The software is able to compare two sequences with partial overlap by means of Bestfit program, contained in GCG package from University of Wisconsin based on the "local homology" algorithm of Smith and Waterman, in order to get the contig. This feature is really relevant when building sequence blocks in large DNA fragment assembly projects .

INTRODUCTION.....	2
1. COMPUTER ANALYSIS OF NUCLEOTIDE SEQUENCES.....	4
Langebecker U., Kleffe J., Borodovsky M. Calculation of characteristics of asymptotic distribution of occurrences of words in genetic linguistics by Marcov chain technique	5
Fedoseyeva V.B., Otenko S.A., Suboch G.M. Nucleotide sequence statistical analysis of pauses in the process of RNA elongation by RNA polymerase of Escherichia coli in vitro	6
Roytberg M.A. Similarity search in two biological sequences	7
Pevzner P.A. Methods of Discrete Mathematics in Molecular Biology	9
Polozov R.V., Tumanyan V.G., Esipova.N.G. Electrostatic of DNA. New dipole model	10
Gelfand M.S. Prediction of protein-coding regions interrupted by introns	11
Schmatchenko V.V., Berezhnev A.B. Mapping of functional regions by graphic representation of long DNA sequences	13
Chuzhanova N.A. The method of recognition of the ribosomal binding sites based on formal grammars	14
Shabalina S.A., Roitberg M.A., Kondrashov A.S., Vernoslov S.E. Some characteristic features of 5'-regulatory regions of heat shock protein genes	16
Leontyev A.Yu. The coloured symmetry of the nucleotide texts and its numerical presentation as a means of nucleotide texts symmetry analysis	18
Volkov V.V., Leontyev A.Yu., Winter V.G. Application of fourier spectroscopy to characterization of origins of replication	20
Shakhmuradov I.A., Gasumov V.A. Search of statistically non-random homologous sites with possible insertions/deletions in the DNA(RNA) sequences	22
Gusev V.D., Kulichkov V.A., Chupakhina O.M. Complexity profiles of genetic texts	24
Alexandrov N.N., Sorokin A.V. Multiple alignment by consensus and recognition problem	26
Mironov A.A. The stochastic properties of Needleman-Wunsch weight matrix	27

Gulyaev A.P., Reshetnikova O.Y., Bechtereva T.A., Pljushin A.Z., Pisareva M.M., Kiselev O.I. The algorithm for search of PCR-primers based on the prediction of the influenza virus RNA secondary structure	28
Pevzner P.A., Kozhukhin C.G. Genome inhomogeneity is determined mainly by WW and SS dinucleotides	29
Mil'shina N.V. The computer prediction of "mRNA stability" signals on 3'-nontranslated regions of nuclear genes	30
Suboch G.M., Sprizhitsky Yu.A. Splice junctions of human genes: statistical analysis and discrimination	32
Monakov J.N., Kisselev O.I. Investigation of homology of nucleotide sequences on basis of the probability conception	33
Monakov J. N., Gulyayev A.P. A dynamic programming algorithm for finding RNA secondary structure with pseudoknots	34
Gulyaev A.P., Monakov J.N. The algorithm and programme for prediction of RNA secondary structure with elements of tertiary interactions based on self-organization principles	35
Konstantinov J.M., Rogozin I.B., Derenko M.V. Homology between plasmid-like DNAs of maize mitochondria and virus sequences examined by contextual analysis	36
Lukina N.I., Soidla T.R. Large stereospecific anomalies of the DNA base composition as an indication for DNA-protein complexes	38
Kapitonov V.V. Phylogenetic analysis in the estimation of transposition frequencies of mobile sequences	40
Ohno S. The grammatical rule for DNA language: messages written in palindromic verses	42
Trifonov E.N., Lagunez J. Translation framing code	44
Berg O.G. Base-pair selection in specific protein binding sites on DNA; a statistical-mechanical model	45
Kleffe J., Langbecker U. Exact computation of pattern probabilities in random sequences generated by Markov chains	46
Bondarenko K.V., Vlaskin N.V., Nizolenko L.Ph., Gusev V.D., Kulichkov V.A. Searching for genetical signals by modified covering method	47
Rogozin I.B., Solovyov V.V., Kolchanov N.A. Method of analysis of oligonucleotide connection with recombinational events	49
Kel' A.E., Ponomarenko M.P., Orlov Yu.L., Ischenko I.V., Kolchanov N.A. The computer system of functional sites analysis in polinucleotide sequences	51

Solovyev V.V., Kel' A.E. New methods for RNA secondary structure calculation. Search for global minimum and accounting of selforganization process	53
Solovyev V.V., Syledtsov I.A., Salikhova A.K. System for revealing of complicated structural-functional organization of genetic texts	55
Rogozin I.B., Solovyov V.V., Sredneva N.E., Kolchanov N.A. Computer system for analysis of molecular mechanisms of point mutation emergence	57
Kolchanov N.A. Systems of artificial intelligence in molecular biology and genetics: problems and perspectives	59
2. COMPUTER ANALYSIS AND MODELLING OF PROTEIN STRUCTURE.....	63
Tumanyan V.G., Rogulenkova V.N., Esipova N.G. The conformation of triple helix of collagen as function of the primary structure	64
Parilis I.I., Kazanov E.Yu. Application of pattern recognition methods for toxin polypeptides identification	65
Petukhov M.G., Dorofeyev V.E., Mazur A.K. Global minimization of the conformational energy of oligopeptides by the tunneling algorithm	67
Dorofeyev V.E., Mazur A.K. Effects of constraints on the dynamics of proteins	69
Demohuk E.J., Tumanyan V.G. Biological adequacy of protein sequence alignments	71
Gorbalenya A.E., Koonin E.V. The universal sequence pattern in a large class of purine NTP-utilizing enzymes	72
Vtyurin N.N. The role of dense packing of hydrophobic groups in β -structural proteins	74
Tseytin V.M., Vakser I.A., Galaktionov S.G. Interaction of peptide and protein molecules with lipophilic environment: computer modelling	75
Zhurkin P.A., Eroshkin A.M., Kulichkov V.A. Computer system for prediction of stabilizing amino acid substitutions in protein. An approach to enhance human gamma-interferon stability	76
Eroshkin A.M., Fomin V.I., Zhilkin P.A., Kulichkov V.A. Protein fragment variability analysis and development of some rules for protein engineering of vaccines	77
Karabut L.V., Peohnyk I.V., Sereyskaya A.A. Modeling of thrombin-fibrinogen interactions connected with a specific recognition	78

Davydov D.R., Fridman G.V., Erlich I.M., Blinov V.M. Profile comparison approach to the prediction of functionally important sites in the proteins. Implementation to the analysis of cytochrome P-450 superfamily	79
Blinov V.M. Prediction of ligand-receptor interaction in virus and cellular proteins	80
Karginov V.A., Lukashev V.A., Blinov V.M. Model of structural organisation of cytochrome P-450 multiple forms	81
Rodionov M.A., Galaktionov S.G. Statistical analysis of residue-residue contacts in globular proteins	82
Andrianov A.M., Sherman S.A. Modelling spatial protein structure by the two-dimensional ¹ H-NMR spectroscopy data	83
Koonin E.V., Chumakov K.M., Gorbalenya A.E., Brodsky L.I., Tatusov R.L. A systematic approach to the localization of conserved motifs in amino acid sequences	85
Benjikh D.N., Ponomarenko M.P., Orlov Yu.L., Kolchanov N.A. The classification of the proteins by indentifying their structural-functional determinants	87
Ponomarenko M.P., Shindyalov I.N. Expert system for protein topology prediction by the amino acid sequences (for personal computer IBM PC)	90
Ponomarenko M.P., Orlov Yu.L. Expert system for fast search of characteristics for biopolymers classification	92
Ponomarenko M.P., Streletz V.B. Expert system for protein sequence functional recognition (for personal computer IBM PC)	94
Smith T.F., Lathrop R., Smith R. The identification of protein functional patterns	96
Bruccoleri R.E., Haber E., Novotny J. Application of conformational search to antibody modeling	98
Pettitt B.M. Prediction of polypeptide secondary structure by molecular integral equations and computer simulations ..	99
Zakrzewska K. Describing protein conformation: a new mathematical approach	100
Crippen G.M. Distance geometry and the calculation of protein conformation	101
Johnson M.S., Overington J., Sali A., Blundell T.L. From the comparative analysis of proteins to knowledge-based modelling	102
Garnier J. Secondary structure prediction and protein design	103

Sklenar H., Eisenhaber F. Inclusion of solvent and counterion effects into the analytical force field of molecular mechanics	104
Polner G. Multiple alignment of protein sequences	106
Bork P., Rohde K., Reich J. Property patterns in protein sequences: automatic recognition and application	107
Eisenmenger F., Sclenar H. Prediction of protein tertiary structure by analogy - modelling of loop regions	108
Tudos E., Cserzo M., Simon I. Protein design using the regularities of protein sequences	109
Tarnowska M., Tempczyk A., Lywo A., Borowski E. Computer modelling of the mechanism of the action of the glucosamine synthase inhibitors	110
Lywo A., Tarnowska M., Grzonka Z., Tempczyk A., Shenderovich M. Use of qsr and molecular mechanics in studying the conformational requirements of the neurohypophyseal hormones ..	111
Solovyov V.V., Salamov A.A. Protein secondary structure calculation on the base of discriminant analysis with the use of information on homologous proteins structure	112
Salamov A.A., Solovyov V.V. Mutations influence on the globular proteins secondary structure	115
Salikhova A.K., Solovyov V.V. α -Helical domains spatial structure calculations on the base of quasispherical primary approximation of protein conformation	117
Benjikh D.N., Ponomarenko M.P., Orlov Yu.L., Kolchanov N.A. The method of computerized presentation of knowledge on proteins in terms of the programs recognizing their structural-functional determinants	120
Benjikh D.N., Salamov A.A., Ponomarenko M.P., Solovyev V.V., Orlov Yu.L., Gushohin D.N., Kolchanov N.A. The reconstruction of the spatial structure of DNA-binding domain of histone H4 by its amino acid sequence	123
Benjikh D.N., Kolchanov N.A., Ponomarenko M.P., Orlov Yu.L. The method of constructing the graph of protein functional similarities on the basis of studying their amino acid sequences	126
Chashchin N.A., Bratus A.V., Maltchenko S.Z. GUHA-method using for the prediction of the protein secondary structure ...	129

3. THE MOLECULAR EVOLUTION THEORY.....	130
Kvinikhidze K.S. Conceptual approach to the analysis of homologous proteins primary structure	131
Shoherbak V.I. The universal and mitochondrial codes: which is older?	132
Serovaiskii S.Y., Birtonov E. Some mathematical aspects of the co-operative symmetry of the genetic code	133
Berman A.L., Dityatev A.E., Frishman D.I. Physico-chemical properties of signal receptor proteins domains as the basis for phylogenetic analysis not requiring sequence alignment	134
Borodovsky M.Yu., Gusein-zade S.M., Mekishev A.A. A ranged distribution of codon frequencies reflects inhomogeneity of DNA's primary-structure evolution	136
Samsonova M.G., Kushnirov V.V., Ter-Avanesyan M.D., Ingevechtomov S.G. Analysis of nucleotide sequence of SUP2 gene (S.Cerevisiae) and it's phylogenetic relation	137
Troitsky A.V., Melekhovets Y., Rakhimova G.M., Bobrova V.K., Valiejo-Roman K.M., Antonov A.S. Phylogenetic trees for several kinds of ribosomal RNA shed light upon early stages of seed plant evolution	139
Krushkal J.S., Rodin S.N. CpG-rich promoters of Alu-like repeats in the mammalian genome: some evolutionary characteristics	141
Matushkin Yu.G., Rodin S.N. Some aspects of coevolutionary complication of genomes	143
Rodin S.N. The concept of coevolution: from molecules to noosphere	144
Matushkin Yu.G. The immanent causes of progressive evolution in the immune system	146
Rzhetsky A.Yu., Rodin S.N., Zharkikh A.A. On molecular recapitulation in globin genes system	148
Rodin S.N., Matushkin Yu.G. The infections striking lymphoid cells take part in the evolution of the molecular-genetic system of immunity	150
Furman D.P., Kozhemyakina T.A., Rodin S.N. On the correlation between penitance and MDG distribution throughout the mutant SCUTE genome in Drosophila melanogaster	152
Ratner V.A. The unified theory of molecular evolution (TME): invitation to discussion	154
Ratner V.A. Genetic language and adjacent problems	156

Kapitonov V.V. Genetic load as a result of Alu-like repeat transposition	158
Kapitonov V.V., Amikishijev V.G. The influence of the "molecular memory" on the intragenome evolution of the mobile genetic elements	160
Kapitonov V.V., Amikishijev V.G. Gene conversion as a factor of preventing regulatory sites from losing validity for MGE	161
Koonin E.V., Gorbalenya A.E. The main linestreams of virus genome evolution	162
Pilyukov A.A. Steady state evolution entropy	164
Hegyi H., Solymosy F. Evolutionary analysis of eight tomato U1 RNA gene candidates	165
Sharp P.M. Codon usage and the silent molecular clock: variation among genes and among organisms	166
Li W.-H., Gouy M. Methods for estimating the number of nucleotide substitutions between sequences under nonequilibrium conditions	168
Bandelt H.-J., Von Haeseler A., Bolick J., Schutte H. A comparative study of sequence dissimilarities and evolutionary distances derived from sets of aligned ribosomal RNA sequences	169
Zharkikh A.A., Rzhetsky A.Yu. The L-plet frequency characteristics of sequences: application to genome evolution analysis	170
Solovyev V.V. Mobile genetic elements role in forming of new patterns of regulatory regions and blok evolution of eukaryotic genome structure	172
Kel' A.E., Kolchanov N.A., Solovyev V.V. Theoretical analysis of the DNA deletion and duplication mechanisms in the prokaryotic genomes on the basis of direct repeats	175
Kel' A.E., Kolchanov N.A., Solovyev V.V. Convergent origin of repeats in genes coding for globular proteins	177
Shindyalov I.N. Computer system for the investigation of molecular evolution on the bases of homological sequences analysis	179
Scudo F.M. A Darwinian theory of molecular evolution	181

4. MOLECULAR-GENETIC SYSTEMS' MODELLING	182
Ratner V.A. Limiting factors and evolution of molecular genetic regulatory systems (MGRS)	183
Blank M.L. Mathematical models of the genetic dynamics of small subdivided populations	184
Dubrova Yu.E. Joint variation of polygenic and monogenic traits in human populations and the problem of optimal level of heterozygosity	186
Likhoshval V.A., Bazhan S.I., Belova O.E. Mathematical modelling of molecular genetic system regulation of interferon induction and antiviral action	187
Abramenko G.G., Chuykov V.V., Kulichkov V.A. An integrated informational system of inflammatory reaction regulation in respiratory virus diseases: Data base and mathematical model	189
Chuykov V.V., Bazhan S.I., Kulichkov V.A. AIDS and AIDS vaccine. Analysis of vaccination perspective for struggle against induced by human immunodeficiency virus (HIV) infection with a help of mathematical model	191
Chuykov V.V., Bazhan S.I., Kulichkov V.A., Buldaev A.S. Mathematical model of antiviral immune response regulation. Investigation of some immunocorrection problems by numerical experiments	193
Prokudina E.I., Tchuraev R.N. Problems of optimal control of bacterial molecular genetics control system dynamics ...	195
Kapitonov V.V., Yashukov V.A., Kolchanov N.A. The populational dynamics of some two- and three-loci molecular-genetic systems	197
Kapitonov V.V., Ponomarenko M.P., Ptitoyn A.A. Relaxation properties of cyclic multy-stage systems	198
Weller K., Petri I., Hartmann M. Organization of gene expression for colony stimulating factor by gene design and RNA secondary structure prediction	200
Lihoshvai V.A. Rare codons: fortuity or regularity?	202
Kel' A.E., Kiseleva E.V., Ischenko I.V., Kolchanov N.A. Mathematical analysis of grand translation polysome's from Chironomus salivary gland cell structural organization	203

5. DATA BANKS AND PROGRAM PACKAGES	205
Golovanov E.I., Subotch G.M. FLEXIS: Systems for management of integrated and inquiry-education databases	206
Akhundova A.A., Lakomova N.M., Jangeldina G.T., Alexandrov A.A. Present state of Genexpress	207
Shoaff W.D., Nuttal T.R., Hozier J.C. A software system for the assembly, manipulation and vusualization of digitally simulated chromosome	208
Grigorjev A.V. Computer simulations of genome mapping by plaque hybridizations	209
Barklaev A., Alexandrov N., Mironov A. The comparison of L-gram and statistical methods for rapid homology search ...	210
Pevzner P.A., Mironov A.A. DNA physical mapping and discrete optimization	211
Shcherbakov V.P., Fonarev A.B., Golovanov E.I. Concepts and structure of organism integrated database	212
Kusmin M.A. Automatic method of visualisation of RNA secondary structures	214
Mil'shina N.V. Sequence signals which may be required for coordinate regulation of genes in mammals: project of database	215
Monakov J.N., Serbin V.M. A package of the computer programs for analysis nucleotide sequences by dynamic programming ..	216
Zharkikh A.A., Rzhetsky A.Yu., Morosov P.S., Sitnikova T.L. VOSTORG: package of a microcomputer programs of phylogenetic analysis	217
Solovyev V.V., Rogozin I.B., Salikhova A.K., Seledtsov I.A., Salamov A.A. "CONTEXT" program package for analysis of structure and function of genetic macromolecules ..	219
Brodsky L.I., Vasiliev N.B., Gorbalenya A.E., Gulyayev A.P., Davydov D.R., Drachev A.L., Koonin E.V., Leontovich A.M., Tatuzov R.L., Chumakov K.M. GENBEE: a package of computer programs for biopolymer sequence analysis	222
Mazanov A.L., Boleznin M.I. X-Site: software package for restriction site identification	224
Ponomarenko M.P., Shindyalov I.N., Streletz V.B. Data bank of biopolymers' three-dimensional structures (for personal computer IBM PC)	225
Kapitonov V.V., Ptitoyn A.A. The package of programs "TRANSREP" for investigating transposable and repetitive sequences	227

Kapitonov V.V., Yashukov V.A. The informational structure of the transposable sequence databases	228
Kohno M.N. MultiScan: software for processing and analysis of electrophoregrams	229
Babski V.G., Zhukov M.Yu. Computer support of electrophoretical experiment	230
Timinskas A.K. New computer programs useful for searching homology patterns and nucleic acids	231
Eroshkin A.M., Zykova N.A., Naumochkin A.N., Kulichkov V.A. Database and information retrieval system on structure-functional organization and activity of peptides and proteins	233
Fomin V.I., Eroshkin A.M., Zhilkin P.A., Niculin A.E. Variable fragments data base and its use in design of chimeric proteins	235
Myasnikova E.M., Knyazev P.G., Borodovsky M.Yu. Software tool for choice of oligonucleotide probe	236
Adamatzky A.I., Ivanov S.F., Bronnikov V.A. A local parallel analysis of nucleotide sequences	237
Babenko V.E., Sagitov V.R. Expert system for genetic engineering experiment planing using quality model	239
Babenko V.E., Sagitov V.R. Vector-PC: The information processing system for genetic engineering	241
Drmanac R., Strezoska Z., Labat I., Radosavljevic D., Paunesku T., Crkvenjakov R. Towards genome DNA sequencing chip based on oligonucleotide hybridization	242
Pevzner P.A. Megabase DNA sequencing and Transportation problem	244
Vassilyev D.G., Korolev S.V., Adzhubei A.A. TEK FRODO: new FRODO version on the base of Tek_MMGP (Tektroniks Molecular Modeling Graphic Package)	246
Mironov A.A., Lunovskaya-Gurova L.V., Bogodarova N.Yu., Alexandrov N.N., Grigorjev A.V., Lebedev V., Pevzner P.A., Trukhan M.E. "DNA-SUN" - a powerfull computer tool for DNA and protein analysis	248
Grigorjev I.V., Grigorjev A.V. A program for cloned DNA analysis	249
Grigorjev A.V. A contig assembly program for genom mapping by plaque hybridizations	250
Solovyev V.V., Salikhova A.K. Data base of DNA, RNA and protein functional regions	251

Makarova K.S., Solovyev V.V., Mazin A.V. "DIROM" - interactive system for oligonucleotide-directed mutagenesis experiments optimization	253
Streletz V.B., Kolchanov N.A. The programming support for the structural-functional analysis of sequences	255
Streletz V.B., Kolchanov N.A. The structural-functional mapping of proteins with the use of the data bank-supplied information on the amino acid sequences	257
Kolchanov N.A., Alexandrov A.A., Kulichkov B.A., Tumanian V.G., Babski V.G., Sprijitsky Yu.A., Kondrashov V.A., Mironov A.A. The basic ideas of the concept of computer-informational investigations of the "HUMAN GENOME" Soviet National Project	259
Sander K. The increasing of the protein database EFF	263
Milanesi L., De Bellis G., Andreoli R. A computer program for automated DNA sequences data handling	264

Подписано к печати 17.07.90 г. МН 02423
Формат 60x90 1/16. Печ. л. 17,2. Уч.-изд.л. 12,5
Тираж 600 экз. Заказ 231 . Цена 95 коп.

Ротапринт Института цитологии и генетики СО АН СССР
630090, Новосибирск, проспект академика Лаврентьева, 10