

## G16B

### BIOINFORMATICS, i.e. INFORMATION AND COMMUNICATION TECHNOLOGY [ICT] SPECIALLY ADAPTED FOR GENETIC OR PROTEIN-RELATED DATA PROCESSING IN COMPUTATIONAL MOLECULAR BIOLOGY

#### Definition statement

*This place covers:*

Methods or systems for genetic or protein-related data processing in computational molecular biology.

Computational processing of data related to nucleic acids, proteins, peptides, or amino acids.

Bioinformatics methods or systems where the digital data processing is inherent or implicit, although not explicitly mentioned.

#### Relationships with other classification places

This subclass covers bioinformatics, whereas subclass [G16C](#) covers computational theoretical chemistry, chemoinformatics and computational materials science.

In order to determine whether classification should be directed to this subclass or to subclass [G16C](#), in particular regarding computational theoretical chemistry ([G16C 10/00](#)) and chemoinformatics ([G16C 20/00](#)), one has to take into account the type of molecule(s), whose characterising features are processed by a computational algorithm.

Following the definition statement, processing of data related to nucleic acids, proteins, peptides and/or amino acids should be classified under [G16B](#).

Processing of data related to any other type of molecule should be classified under [G16C](#).

#### References

##### Informative references

*Attention is drawn to the following places, which may be of interest for search:*

Medical diagnosis	<a href="#">A61B 5/00</a>
Genetic engineering involving nucleic acids	<a href="#">C12N 15/00</a>
Nucleic acid analysis, e.g. microarrays, sequencing or PCR	<a href="#">C12Q 1/68</a>
Chromatographic signal analysis	<a href="#">G01N 30/86</a>
Chemical analysis of biological material, e.g. blood or urine	<a href="#">G01N 33/50</a>
Chemical analysis of biological material involving proteins, peptides or amino acids	<a href="#">G01N 33/68</a>
Computer input/output arrangements	<a href="#">G06F 3/00</a>
Computer architectures or program control	<a href="#">G06F 9/00</a>
Information retrieval; Database structures thereof; File system structures therefor	<a href="#">G06F 16/00</a>
Complex mathematical operations	<a href="#">G06F 17/10</a>
Pattern recognition	<a href="#">G06F 18/00</a>
Computer systems using neural network models per se	<a href="#">G06N 3/02</a>
Computer systems using knowledge representation per se, e.g. expert systems	<a href="#">G06N 5/02</a>

Computer systems using probabilistic models per se	<a href="#">G06N 7/00</a>
Machine learning	<a href="#">G06N 20/00</a>
3D image rendering	<a href="#">G06T 15/00</a>
3D modelling for computer graphics	<a href="#">G06T 17/00</a>
Manipulating 3D models or images for computer graphics	<a href="#">G06T 19/00</a>
Computational chemistry; Chemoinformatics	<a href="#">G16C</a>
Computational materials science	<a href="#">G16C 60/00</a>
Healthcare Informatics	<a href="#">G16H</a>
Mass spectrometry apparatus per se	<a href="#">H01J 49/00</a>

## Glossary of terms

*In this place, the following terms or expressions are used with the meaning indicated:*

systems biology	Simulation and mathematical modelling of relationships and interactions between molecular entities in sub-cellular systems integrating genetic and/or protein-related data to describe the dynamic behaviour of, for example, protein-protein/protein-ligand interactions, regulatory networks and metabolic networks
phylogeny	Reconstruction of an evolutionary development and history of a species or higher taxonomic grouping of organisms; typically represented as a phylogenetic tree; methods for creating phylogenetic trees
phylogenetic tree	Tree-like graphical representation of phylogenetic relationships
molecular structure	2-dimensional or 3-dimensional arrangement of atoms, groups of atoms or domains in nucleic acids, proteins, peptides and amino acids
structure alignment	Form of alignment to establish structural and functional equivalences between two or more proteins based on their secondary or tertiary structures
protein folding	Process by which a polypeptide chain folds into a specific 3-dimensional structure
domain	Domain of a protein is an element of the overall molecular structure that is self-stabilising and often folds independently of the rest of a polypeptide chain
drug targeting	Drug design strategy aiming at optimising the properties of a medicinal compound, based on the 3-dimensional structure of a target, for delivery to a particular tissue or organ in the body
functional genomics	Experimental analyses aiming at assessing the function of genes in determining traits, physiology and/or development of an organism, making use of computational and high-throughput technologies
proteomics	Large-scale study of the functions of proteins and their interactions with other molecular entities in a biological system
genotype	Genetic makeup or profile of an organism with respect to a trait
ploidy	Number of sets of chromosomes in a cell/cells of an organism
allele	Alternative form of a gene (one member of a pair) that is located at a specific position (locus) on a specific chromosome

snp	Single nucleotide polymorphism: a DNA sequence variation that involves a change in a single nucleotide and is commonly present in a part of a population
motif	Specific nucleotide or amino acid sequence pattern
population genetics	Study of genetic variation and genetic evolution of populations
linkage disequilibrium	Tendency of alleles located close to each other on the same chromosome to be inherited together
mutagenesis	Process by which the genetic information of an organism is changed, resulting in a mutation
gene expression	Process by which proteins are made or transcribed from the instructions encoded in DNA
gene expression profiling	Determination of the pattern of genes expressed, i.e. transcribed, under specific circumstances or in a specific cell line
probe design and optimisation for microarrays	Designing and selecting (i) optimal, highly specific probes, e.g. oligonucleotides, cDNA, fragments for hybridisation experiments with microarrays and (ii) optimal sets of probes, e.g. oligonucleotides, cDNA, to be chemically attached to a solid support to form an array
microarray	Plurality of nucleic acid probes attached to a substrate, which form an ordered pattern
sequence alignment	Process of comparing nucleic or amino acid sequences, generally by a linear alignment in such a way that equivalent positions in adjacent sequences are brought into the correct alignment with each other by introducing insertions in suitable positions, in order to identify similarities and/or differences amongst the compared sequences
sequence assembly	Method by which linear portions of sequence information are assembled to obtain full length gene sequence data
in silico	Performed on a computer or via computer simulations
ontology	Classification methodology for formalising a subject's knowledge in a structured and controlled vocabulary

### Synonyms and Keywords

*In patent documents, the following words/expressions are often used with the meaning indicated:*

systems	includes apparatus
---------	--------------------

## G16B 5/00

**ICT specially adapted for modelling or simulations in systems biology, e.g. gene-regulatory networks, protein interaction networks or metabolic networks**

### Definition statement

*This place covers:*

Simulation or mathematical modelling of relationships and interactions between molecular entities on a subcellular level, integrating data related to genetic and/or proteins to describe the dynamic behaviour of protein-protein/protein-ligand interactions, regulatory or metabolic networks.

Mere mention of modelling or simulation is not sufficient to justify classification in this place.

**G16B 10/00****ICT specially adapted for evolutionary bioinformatics, e.g. phylogenetic tree construction or analysis****Definition statement**

*This place covers:*

Analysis of orthologous, paralogous, syntenic, or taxonomic relationships.

Generation of pedigrees and phylogenetic trees.

Mere mention of evolutionary data is not sufficient to justify classification in this place.

**G16B 15/00****ICT specially adapted for analysing two-dimensional or three-dimensional molecular structures, e.g. structural or functional relations or structure alignment****Definition statement**

*This place covers:*

Structural architecture of proteins, peptides, amino acids, and nucleic acids and the prediction thereof.

Processes including structural alignment, protein folding, domain topology, molecular modelling, receptor-ligand modelling, docking methods, structural-functional relationships and drug targeting using structure data, as well as two- and three-dimensional structure prediction and/or analysis.

The structure types include secondary, tertiary, and quaternary structures.

Mere mention of structural data is not sufficient to justify classification in this place.

**G16B 20/00****ICT specially adapted for functional genomics or proteomics, e.g. genotype-phenotype associations****Definition statement**

*This place covers:*

Assessment of the function of genes and proteins in determining traits, physiology and/or development of an organism, making use of computational and large scale, high-throughput technologies.

Genotypic-phenotypic associations, including genotyping and genome annotation, linkage disequilibrium analysis and association studies, population genetics, alternative splicing, and small Interfering RNA design (siRNA, RNAi).

Binding site identification, mutagenesis analysis, protein-protein or protein-nucleic acid interactions.

Mere mention of gene or protein function is not sufficient to justify classification in this place.

## **G16B 25/00**

### **ICT specially adapted for hybridisation; ICT specially adapted for gene or protein expression**

#### **Definition statement**

*This place covers:*

Analysis of hybridization or gene/protein expression information. This includes microarray analysis, gel electrophoresis analysis and sequencing by hybridisation (SBH). Further covered technologies include modelling polymerase chain reaction (PCR) data, primer or probe design and probe optimisation, microarray design, normalisation, expression profiling, noise correction models, and expression-ratio estimation.

Mere mention of hybridisation or gene/protein expression is not sufficient to justify classification in this place.

## **G16B 30/00**

### **ICT specially adapted for sequence analysis involving nucleotides or amino acids**

#### **Definition statement**

*This place covers:*

Comparison of sequence information, wherein the sequences are nucleic acids, proteins, or peptides. The comparisons include methods of alignment, homology identification, motif identification, single-nucleotide polymorphism (SNP) discovery, haplotype identification, fragment assembly, and gene finding.

Mere mention of sequence data is not sufficient to justify classification in this place.

## **G16B 35/00**

### **ICT specially adapted for in silico combinatorial libraries of nucleic acids, proteins or peptides**

#### **Definition statement**

*This place covers:*

In silico (i.e. computer based) designing and screening of combinatorial nucleic acids, protein, or peptide libraries.

Mere mention of nucleic acid, protein, or peptide combinatorial libraries is not sufficient to justify classification in this place.

## **G16B 40/00**

**ICT specially adapted for biostatistics; ICT specially adapted for bioinformatics-related machine learning or data mining, e.g. knowledge discovery or pattern finding**

### **Definition statement**

*This place covers:*

Discovery and/or analysis of patterns within a vast amount of genetic or protein-related data, wherein the emphasis is placed on the method of analysis and is largely independent of the particular type of bioinformatic data.

Covered methods based on machine learning and statistical models; supervised and unsupervised learning techniques include bioinformatic pattern finding, knowledge discovery, rule extraction, correlation, clustering, and classification.

Multivariate analysis of protein or gene-related data, e.g. analysis of variances (ANOVA), principal component analysis (PCA), and support vector machines (SVM).

## **G16B 45/00**

**ICT specially adapted for bioinformatics-related data visualisation, e.g. displaying of maps or networks**

### **Definition statement**

*This place covers:*

Visual representations specifically adapted to bioinformatic data, wherein the emphasis is placed on the method of visualisation and is largely independent of the particular type of bioinformatic data.

For example: graphics generation, map display (e.g. haplotype maps, linkage maps), and network display (e.g. genetic networks, protein-protein interaction networks, metabolic networks).

## **G16B 50/00**

**ICT programming tools or database systems specially adapted for bioinformatics**

### **Definition statement**

*This place covers:*

Software specially adapted to assist in programming procedures within bioinformatics.

Database systems specially adapted for managing bioinformatic data. For example: ontologies, heterogeneous data integration, data warehousing, and computing architectures.

Encryption and compression algorithms for genetic data.