TEACHING PROTEIN SCIENCE WITH APPLICATION OF COMPUTERS – CASE STUDY OF BCL-2 PROTEIN FAMILY

Agnieszka Kaczor, Dariusz Matosiuk
Medical University of Lublin, Poland
E-mail: agnieszka.kaczor@am.lublin.pl

Andrzej Persona
Maria Curie-Skłodowska University, Lublin, Poland
E-mail: persona@hermes.umcs.lublin.pl

Abstract

Proteins constitute the main group of molecular targets for most approved drugs. Moreover, more than 50% of drugs target proteins are derived from only four gene families: class I G protein-coupled receptors (GPCRs), nuclear receptors, ligand-gated ion channels and voltage-gated ion channels. The above facts make the prediction of protein structures and the design of ligands which modulate their activity an essential skill for a contemporary medicinal chemist. However, both medicinal chemistry and bioinformatics are often neglected in circulars for chemistry and pharmacy studies.

This article presents a proposal of computer classes on protein structure and function (10 hours), designated for the students of the third or fourth year of biology, chemistry or pharmacy. The course gets the students familiar with protein databases (eg. SwissProt and Protein Data Bank) as well as typical software for visualization of protein structures (eg. SPDBV, PyMol, Yasara). The participants get to know the diversity of proteins (GPCRs, ion channels, enzymes), learn how to analyze different levels of protein structures, to make a sequence alignment, to create a simple homology model and to analyze ligand-protein interactions. Most practical exercises concern the family of BCL-2 proteins which are apoptosis regulators.

Key words*: protein science, proteins as drug targets, BCL-2 protein family.*

Introduction

Genomics and proteomics technologies have led to a paradigm shift in the drug discovery process (Jiang & Zhou, 2005). Bioinformatics has now a key role in the exploitation of genomic, transcriptomic, and proteomic data to investigate the molecular mechanisms of diseases and to identify potential drug targets. Application of computer-aided drug design (CADD) techniques to the drug discovery approaches may lead to a reduction of up to 50% in the cost of drug design (Taft, da Silva & da Silva, 2008).

Proteins constitute the main group of molecular targets for the most approved drugs. Furthermore, more than 50% of drug target proteins are derived from only four gene families: class I G protein-coupled receptors (GPCRs), nuclear receptors, ligand-gated ion channels and voltage-gated

ion channels. Except for nuclear receptors, the main drug targets are membrane proteins. Due to the fact that the detailed 3D structures for the majority of membrane proteins are still unknown, construction of protein models is a useful alternative for investigation of ligand-protein interaction.

The above facts make the prediction of protein structures and the design of ligands which modulate their activity an essential skill for a contemporary medicinal chemist. However, bioinformatics and medicinal chemistry are often neglected in circulars for chemistry and pharmacy studies.

The aim of this article is to present a proposal of computer classes on protein structure and function. The classes involve 10 hours of teaching (three meetings, 3 or 4 hours each) and are designated for the students of the third or fourth year of biology, chemistry or pharmacy. The subjects of the meetings are: 1. Introduction to protein science; 2. Investigation of primary and secondary structure of proteins; 3. Construction of 3D protein models. Most practical exercises concern the family of BCL-2 proteins which are apoptosis regulators.

Results

Meeting 1: Introduction to protein science

The purpose of the classes is to present diversity of protein structures and functions and to acquaint students with protein databases and typical software used for visualization of protein structures.

In the introductory part of the classes the students recall the basic information about proteins concerning 20 standard amino acids and their properties as well as levels of protein structure. Using the material which they obtained before the meeting, the students discuss diversity of function of proteins: binding of other molecules, catalysis, switching and structural functions (compare Table 1, elaborated on the base of data by Petsko & Ringe 2004).

Table 1. Examples of biochemical functions of proteins.

Function	Examples	Description
Binding	TATA binding protein (PDB code 1TGH)	The TATA binding protein binds a specific DNA sequence and serves as the platform for the complex which initiates transcription of genetic information
	Myoglobin (PDB code 1A6K)	Myoglobin binds a molecule of oxygen in a reversible way to the iron atom in its heme group. It stores oxygen for use in muscle tissue.
Catalysis	DNA polymerase (PDB code 1 PBX)	DNA polymeraze is responsible for the catalysis of DNA replication. It copies the genetic material and edits the product for the errors in the copy.
	HIV protease (PDB code 1A8K)	Replication of the AIDS virus HIV depends on the action of protein-cleaving enzyme called HIV protease.
Switching	GTPase Ras (PDB code 1PLL)	The conformational change that occurs in the GTPase Ras is critically important for the molecular basis of many types of cancer.
Structural proteins	Silk (PDB code 1SLK)	Silk is a giant stack of antiparallel beta sheets. Its strength derives from covalent and hydrogen bonds within each sheet, whereas the flexibility results from van der Waals interaction which hold the sheets together.
	F-actin (PDB code 2ZWH)	Actin fibres are important for muscle contraction and for the cytoskeleton.

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Volume 11, 2009

The next part of the classes is devoted to protein databases: SwissProt database of protein sequences and Protein Data Bank of Protein structures. The students are encouraged to find any protein of their interests in both databases and to discuss information which can be found in a protein record. Furthermore, they are asked to find the following proteins in Protein Data Bank: potassium channel (PDB code 1J95), β2 adrenergic receptor (PDB code 2RH1) (compare Figure 1), monoamine oxidase B (MAO-B) in complex with benzylhydrazine (PDB code 2VRL) (compare Figure 2). The students inspect the above proteins with application of SPDBV (Guex & Peitsch, 1997), PyMol (De Lano, 2002) and Yasara View (Krieger & Vriend, 2002). They try different rendering methods and identify elements of secondary structure (helices, β-sheets, coils). They analyze the ligand-enzyme interactions in the MAO-B-benzylhydrazine complex.

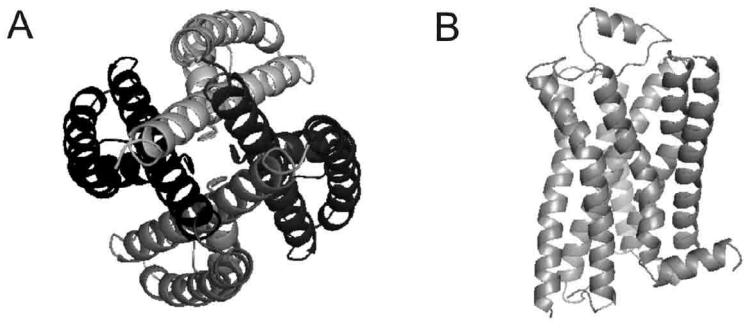

Figure 1. **3D structure of typical membrane proteins. A – KCSA potassium channel; B – β2 adrenergic receptor.**

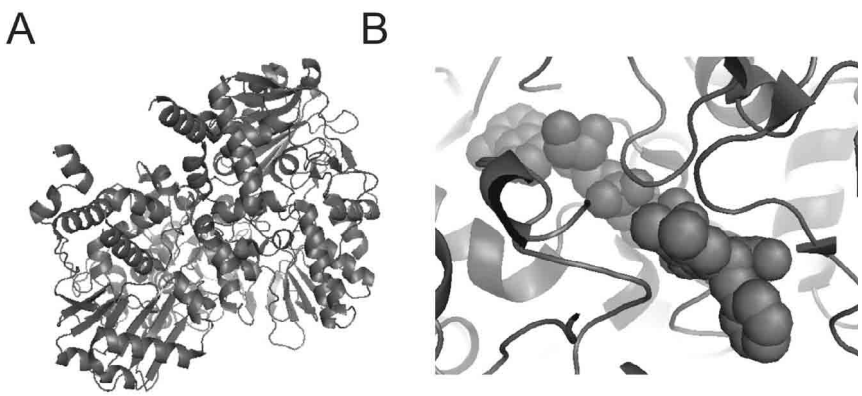

Figure 2. **A – a dimer of monoamine oxidase B; B – ligand in the binding site of MAO-B.**

Meeting 2: Investigation of primary and secondary structure of proteins

The purpose of the meeting is to introduce students to the structure and functions of proteins from BCL-2 family and to demonstrate possibilities of web-based tools for investigation of primary and secondary structures of proteins.

BCL-2 protein is the prototype of a family of proteins containing at least one BCL-2 homol-

ogy (BH) region. In humans, the BCL-2 family is divided into anti-apoptotic multi-domain proteins (prototypes: BCL-2 and BCL-XL), which contain four BH domains (numbered BH1 to BH4), pro-apoptotic multi-domain proteins (prototypes: BAX and BAK), which contain three BH domains (BH1, BH2 and BH3), and the pro-apoptotic BH3-only protein family (which has more than a dozen members) (Levine, Sinha & Kroemer, 2008). Overexpression of members of the BCL-2 family of anti-apoptotic proteins is commonly associated with unfavorable pathogenesis in cancer (Lessene, Czabotar & Kolman, 2008).

The first part of the classes is devoted to analysis of the SwissProt record of human BCL-2 protein (accession number P10415). The students discuss the number of residues, protein motifs and variants as well as elements of secondary structure as they are listed in the database. They follow the "Similarity" link to access the list of other members of BCL-2 protein family. Finally, they look for BCL-2 protein structure in Protein Data Bank (PDB code 1GJH) and investigate it with the earlier introduced software (compare Figure 3).

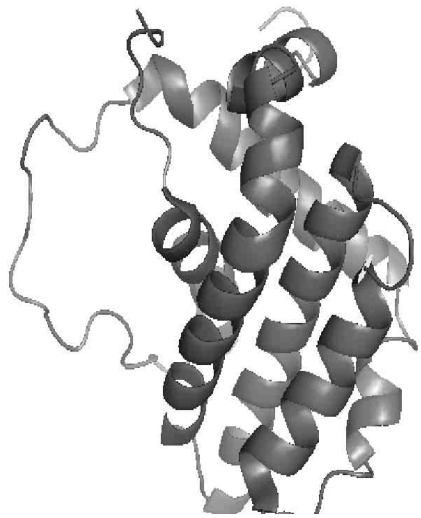

Figure 3. 3D structure of human BCL-2-protein with putative flexible loop replaced with a portion of apoptosis regulator BCL-X protein.

The second part of the classes concerns the analysis of primary and secondary structure of human BCL-2 protein. The analysis of primary structure is performed with application of ProtParam software (Gasteiger et al., 2005), available at EXPASY server. The results of investigation are listed in Tables 2-3. The secondary structure of BCl-2 protein, predicted with PSIPRED (Jones, 1999), is presented in Figure 4. The predicted ranges of helices are compared with the exact helical regions from a suitable PDB file.

Table 2. Parameters of human BCl-2 protein calculated with ProtParam software.

Parameter	Value
Molecular formula	$C_{1186}H_{1800}N_{330}O_{331}S_9$
Number of amino acids	239
Theoretical pI	6.75
Total number of negatively charged residues (Asp + Glu)	22
Total number of positively charged residues (Arg + Lys)	21

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Volume 11, 2009

Parameter	Value
Atomic composition	Carbon ,C 1186 Hydrogen, H 1800 Nitrogen, N 330 Oxygen, O 331 Sulfur, S 9
Estimated half-life	The N-terminal of the sequence considered is M (Met). The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro); >20 hours (yeast, in vivo); >10 hours (*Escherichia coli*, in vivo).
Instability index	The instability index (II) is computed to be 51.63. This classifies the protein as unstable.
Aliphatic index	78.03
Grand average of hydropathicity (GRAVY)	-0.136

Table 3. Amino acid composition of human BCl-2 protein calculated with ProtParam software.

Amino acid	Number	Percentage
Ala	27	11.3
Arg	17	7.1
Asn	6	2.5
Asp	12	5.0
Cys	2	0.8
Gln	6	2.5
Glu	10	4.2
Gly	22	9.2
His	9	3.8
Ile	7	2.9
Leu	22	9.2
Lys	4	1.7
Met	7	2.9
Phe	12	5.0
Pro	19	7.9
Ser	15	6.3
Thr	12	5.0
Trp	6	2.5
Tyr	8	3.3
Val	1	6.7

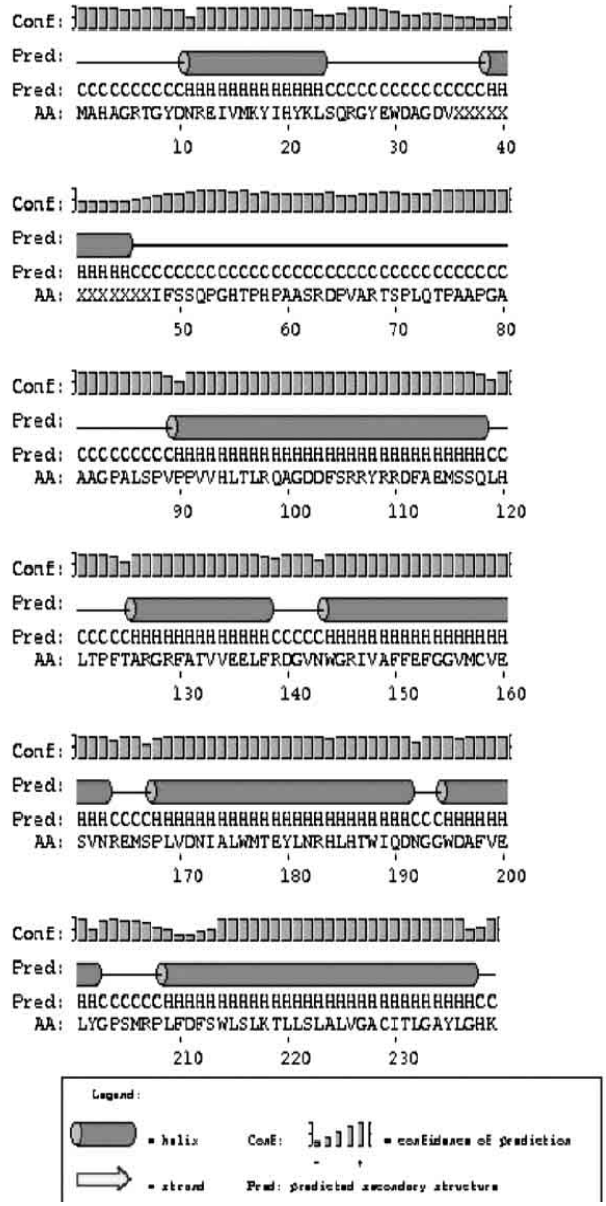

Figure 4. Prediction of secondary structure with PSIPRED for human BCL-2 protein.

Meeting 3: Construction of 3D protein models

The purpose of the classes is to introduce students to multiple sequence alignments and automatic generation of homology models.

First, students are asked to perform multiple sequence alignment for several members of BCL-2 protein family with T-COFFEE software (Notredame, Higgins & Heringa, 2000) (available at EXPASY server) and to identify evolutionary conserved residues (compare Figure 5). Then, they construct a homology model of human BOK protein (BCL-2-related ovarian killer protein, accession number Q9UMX3) with SWISSMODEL (Arnold, Bordoli, Kopp & Schwede, 2006) (compare Figure

6). BCL-W protein (PDB code 1ZY3) was selected as a template. The sequence identity between the target and the template was 19.412. Finally, the students analyze the quality of the obtained model with VERIFY3D (Eisenberg, Lüthy & Bowie, 1997) (compare Figure 7).

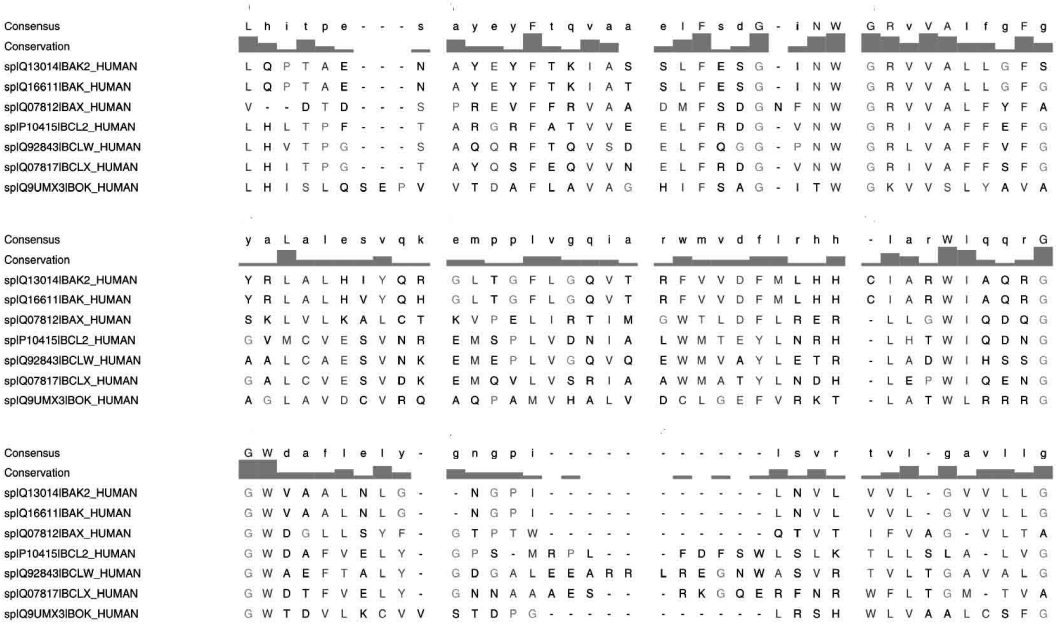

Figure 5. The fragment of multiple sequence alignment of BCL-2 protein family. Visualization with Chimera (Petteresen et al. 2004).

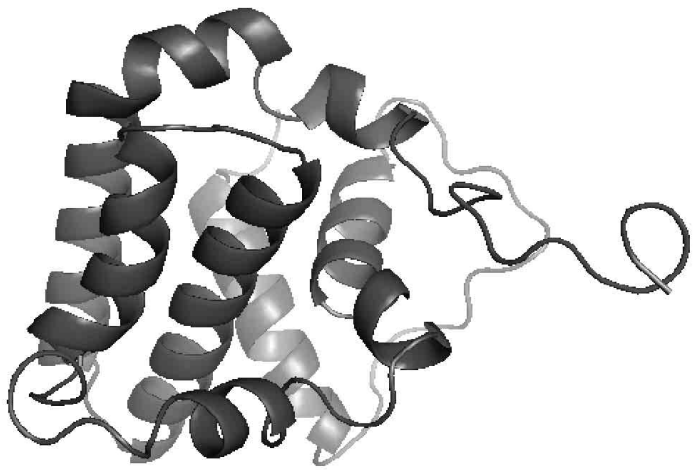

Figure 6. The model of human BOK protein obtained with SWISSMODEL.

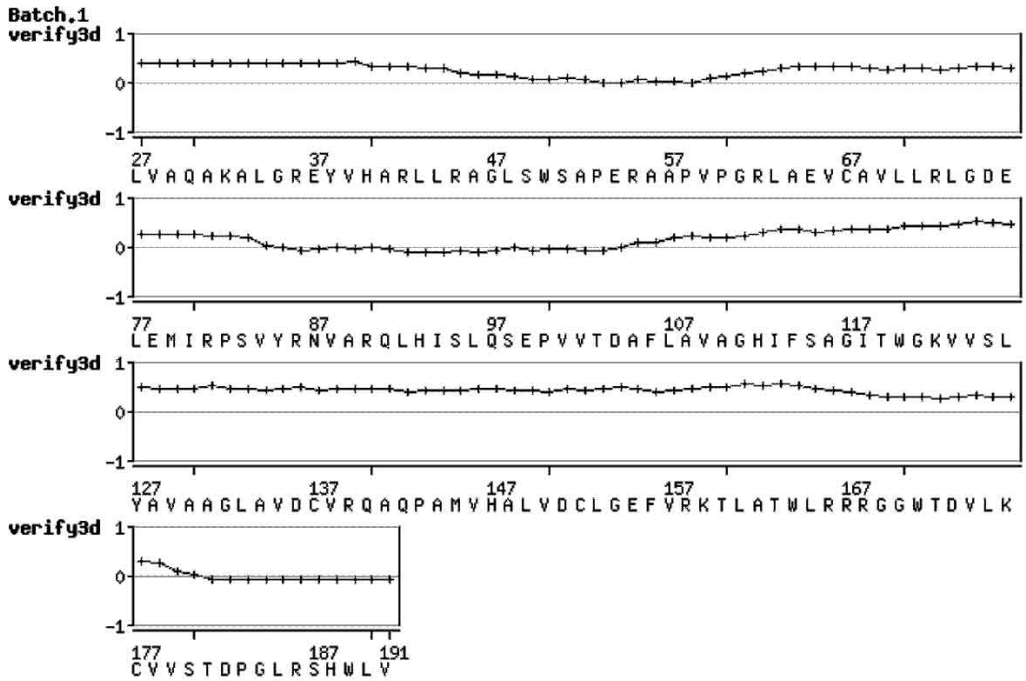

Figure 7. Estimation of the model of human BOK protein with VERIFY3D. The scores range from -1 (bad score) to +1 (good score).

Discussion

The presented proposal of 10-hour classes on protein structure and function may be treated as short and quick introduction to protein science. The course gets the students familiar with protein databases (eg. UniProt and Protein Data Bank) as well as typical software for visualization of protein structures (eg. SPDBV, PyMol, Yasara). The participants get to know the diversity of proteins (GPCRs, ion channels, enzymes), learn how to analyze different levels of protein structures, to make a sequence alignment, to create a simple homology model and to analyze ligand-protein interactions. In spite of their time limitations, such classes performed for the fourth year pharmacy students at Medical University of Lublin met with enthusiastic perception of participants. The attending students emphasized that for most of them it was the first time when they saw a three-dimensional structure of any molecule. The classes changed student notion about membrane receptors. The participants, who are on the threshold of their future career selection, gain a foretaste of a medicinal chemist's job at a university or in pharmaceutical industry. It is important, however, to inform students that real case studies on protein modeling are much more complex and require more advanced software. It is worth stressing that it did not discourage at least some of them as several participants of the classes chose protein modeling as the subject of their master thesis.

The proposed classes require only freely available software or publicly accessible web servers so they generate no costs with software licenses.

Conclusions

The presented computer classes on protein structure and function may be utilize in education of pharmacy and chemistry but also biology and medicine students.

During the 10-hour course the students get familiar with the basics of protein modeling: investigation of protein primary structure, prediction of the secondary structure, performing of sequence alignment and construction of simple homology models.

Most practical exercises are based on the attractive molecular target, BCL-2 protein family, weakly known earlier by most participants.

The proposed classes require freely available software and web servers only.

References

Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22,195-201.

De Lano, W.L. (2002). *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos CA, USA.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. et al. (2005). Protein identification and analysis tools on the ExPASy Server [in:] John M. Walker (ed): *The proteomics protocols handbook*, Humana Press.

Jiang, Z. & Zhou, Y. (2005). Using bioinformatics for drug target identification from the genome. *American Journal of Pharmacogenomics*, 5, 387-396.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices, *Journal of Molecular Biology*, 292, 195-202.

Krieger, E. & Vriend, G. (2002). Models@Home - Distributed computing in bioinformatics using a screensaver based approach. *Bioinformatics* 18, 315-318

Lessene, G., Czabotar, P.E. & Colman, P.M. (2008). BCL-2 family antagonists for cancer therapy. *Nature Reviews Drug Discovery*, 7, 989-1000.

Levine, B., Sinha, S. & Kroemer, G. (2008). BCL-2 family members: dual regulators of apoptosis and autophagy. *Autophagy* 4, 600-606.

Notredame, C., Higgins, D.G. & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, 302, 205-217.

Petsko, G.A. & Ringe, D. (2004). *Protein Structure and Function*, New Science Press Ltd.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. et al. (2004). UCSF Chimera - a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25, 1605-1612

Taft, C.A., Da Silva, V.B. & Da Silva, C.H. (2008). Current topics in computer-aided drug design. *Journal of Pharmaceutical Science*, 97, 1089-1098.

Web addresses: Expasy Proteomics Server: http://www.expasy.ch; Protein Data Bank: http://www.rcsb.org

Adviced by Jolanta Kotlinska,
Medical University of Lublin, Poland

Agnieszka Kaczor	Assistant researcher, post-doc, Medical University of Lublin, Faculty of Pharmacy, Department of Synthesis and Chemical Technology of Pharmaceutical Substances, 6 Staszica Street, 20-081 Lublin, Poland. E-mail: agnieszka.kaczor@am.lublin.pl Website: http://www.am.lublin.pl
Dariusz Matosiuk	Head of Department, Medical University of Lublin, Faculty of Pharmacy, Department of Synthesis and Chemical Technology of Pharmaceutical Substances, 6 Staszica Street, 20-081 Lublin, Poland, E-mail: darek.matosiuk@am.lublin.pl Website: http://www.am.lublin.pl
Andrzej Persona	Assistant researcher, Maria Curie-Sklodowska University, Faculty of Chemistry, Department of Analytical Chemistry and Instrumental Analysis, pl. M. Curie-Sklodowskiej 3, 20-031 Lublin, Poland. E-mail: persona@hermes.umcs.lublin.pl Website: http://www.umcs.lublin.pl/