EDITORIAL BOARD

Polish Ministry of Science and Higher Education journal rating: 6.000
Sustaining institution: Ministry of Science and Higher Education
Edition: 300 copies

ISSN 1895-9091 (hard copy)
ISSN 1896-530X (electronic version)

http://www.bams.cm-uj.krakow.pl

The publisher: Księgarnia Akademicka, Kraków, St. Anny 6, Poland
http://www.akademicka.pl

Contents

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 5-12

AB INITIO PROTEIN STRUCTURE PREDICTION – THE HYDROPHOBICITY DISTRIBUTION ANALYSIS

Giuseppe Evangelista[1], Giovanni Minervini[1], Fabio Polticelli[1], Maciej Malawski[2,3], Tomasz Szepieniec[3], Lukasz Flis[3], Katarzyna Prymula[4,5], Marek Kochanczyk[4,6], Ewa Matczynska[4,7], Zdzislaw Wisniowski[4], Kinga Salapa[4], Mateusz Banach[4,6], Irena Roterman[4]

[1] Department of Biology, University Roma Tre, Viale G. Marconi 446, I-00146 Rome
[2] Institute of Computer Science AGH, Mickiewicza 30, PL-30-059 Krakow
[3] Academic Computer Center CYFRONET, Nawojki 11, PL-30-0950 Krakow
[4] Department of Bioinformatics and Telemedicine, Lazarza 16, PL-31-530 Krakow
[5] Faculty of Chemistry, Jagiellonian University, Ingardena 3, PL-30-060 Krakow
[6] Faculty of Physics, Astronomy and Applied Computer Science – Jagiellonian University, Reymonta 4, PL30-059 Krakow

Abstract: The three-dimensional structures generated for 20 "never born proteins" (NBP – random amino acid sequence with no significant homology to existing proteins) using two different techniques: ROSETTA (called R in the paper) and "fuzzy oil drop" model (called S in the paper) were compared to estimate the accordance with the assumed model estimating the influence of an external force field on the final structure of the protein. Selected structures are those corresponding to the highest (10 proteins) and lowest (10 proteins) RMS-D values obtained measuring the similarity between the R and S structures. The R structures generated according to an internal force field (the individual inter-molecular interaction) including solvation effects were analyzed using the "fuzzy oil drop" model as target model. The second applied model "fuzzy oil drop" generated structures characterized by an ordered hydrophobic core structure. 13 of the 20 selected S structures appeared to be accordant with the "fuzzy oil drop" model while 6 out of the 20 structures appeared to be accordant with external force field for R structures which suggests a general interpretation of the influence of an external force field on the folding simulation.

Keywords: hydrophobicity, protein structure, oil drop model

Introduction

Protein structure prediction is a critical problem in bioinformatics which is so far unable to deliver the tools to reliably predict the structure of a protein from its amino acid sequence (the complete information available on webpage [1]. On the other hand the huge number of protein sequences theoretically possible which do not show any significant similarity with real proteins seems to hide a large number of possible biological activities of pharmacological significance [2, 3], a problem which cannot be approached using experimental methods

From this viewpoint, one of the software tools which in the recent years showed a significant success is the ROSETTA ab initio program developed by Baker and coworkers [4]. Rosetta *abinitio* is an *ab initio* protein structure prediction software which is based on the assumption that in a polypeptide chain local interactions bias the conformation of sequence fragments, while global interactions determine the three-dimensional structure with minimal energy which is also compatible with the local biases. The energy minimization procedure applied by Rosetta on the models generated uses a semi-empirical force field in which the principal non-local interactions considered are hydrophobic interactions, electrostatic interactions, main chain hydrogen bonds, solvation energy and excluded volume [4].The internal energy expressing the pair-wise interactions in a protein molecule with pair-wise interaction with surrounding water molecules generates the structure with a limited influence of external force field (structures called R in this paper). The alternate model taking into account the influence of external force field applied in "fuzzy oil drop" model (structures called S in this paper) is aimed at revealing the influence of an external force field on the final structure resulting from the simulation.

Materials and methods

Sequences of Never Born proteins

The 70 amino acids long random amino acid sequences of NBPs were generated using a previously described RandomBLAST software tool [5]. RandomBLAST generates sequences with equal amino acid frequency for all the 20 amino acids and discards sequences which display significant similarity with known natural protein sequences.

Structures

A total of 10 000 NBPs modeled structures were generated both using ROSETTA *abinitio* software [6] and the "fuzzy oil drop" model [7]. The ROSETTA *abinitio* method is described in detail in [4]. The structures generated by ROSETTA *abinitio* are called R in this paper.

The "fuzzy oil drop" model based on the assumption of the hydrophobic core (represented by 3-D Gauss function) generation was described in details in [5]. The structures generated according to "fuzzy oil drop" model are called S in this paper.

The R and S structures were compared using the standard RMS-D calculation.

The ten structures corresponding to the lowest RMS-D values calculated comparing the R and S forms and the ten structures corresponding to the highest RMS-D values were selected for analysis [8]. Both categories of structures (R and S) were compared using the hydrophobicity density distribution along the polypeptide chain.

Comparison of hydrophobicity distributions

Theoretical distribution

The idealized hydrophobicity distribution is assumed to be expressed by 3-D Gauss function: the concentration reaching zero on the surface of the "drop", according to the Gauss function:

$$\tilde{H}t_j = \frac{1}{\tilde{H}t_{sum}} \exp\left(\frac{-(x_j-\bar{x})^2}{2\sigma_x^2}\right)\exp\left(\frac{-(y_j-\bar{y})^2}{2\sigma_y^2}\right)\exp\left(\frac{-(z_j-\bar{z})^2}{2\sigma_y^2}\right), \quad (1)$$

where x̄, ȳ, z̄ are the coordinates of the geometric center of the molecule (usually located in the origin of the coordinate system). This is why these values can be considered equal to zero. The size of the molecule is expressed by the triple σ_x, σ_y, σ_z, which is calculated for each molecule individually provided that the orientation of the molecule with the longest possible inter-effective atoms distance is determined according to the appropriate coordinate system axis. The σ values are calculated as the 1/3 of the longest distance between two effective atoms calculated along each axis. The value of the Gauss function at any point of protein body is treated as the idealized hydrophobic density defining the hydrophobic core.

The idealized hydrophobicity at any point of the "fuzzy oil drop" can be calculated according to the Gauss function for the molecule located with its geometric center as the origin of the coordinate system. On the other hand, the empirical hydrophobicity distribution is calculated according to the function presented by Levitt [9]:

$$\tilde{H}o_j = \frac{1}{\tilde{H}o_{sum}} \sum_{i=1}^{N} (H_i^r + H_j^r) \begin{cases} \left[1 - \frac{1}{2}\left(7\left(\frac{r_j}{c}\right)^2 - 9\left(\frac{r_j}{c}\right)^4 + 5\left(\frac{r_j}{c}\right)^6 + \left(\frac{r_j}{c}\right)^8\right)\right] \\ 0 \text{ for } r_j > c \end{cases} \Bigg] 0 \text{ for } r_j \le c \quad (2)$$

where *N* expresses the number of amino acids in the protein (number of grid points), $\tilde{H}_i^r$ expresses the hydrophobicity of the *i-th* residue according to the accepted hydrophobicity scale. The values of $\tilde{H}o_j$ are standardized by dividing them by the coefficient $\tilde{H}o_{sum}$, which is the sum of all hydrophobicities attributed to grid points.

The $\Delta\tilde{H}$ values express the summary of differences between the expected ($\tilde{H}_T$) hydrophobicity distribution and the one observed ($\tilde{H}_O$) in a particular protein calculated for each residue individually.

$$\Delta\tilde{H} = \tilde{H}_{Ti} - \tilde{H}_{Oi} \quad (3)$$

The idealized hydrophobicity density distributions (T) (according to a 3-D Gauss function) as well as the observed ones (O) in final structures R and S were used for entropy calculation. The idealized (T) and random (Rd) distributions were taken as the reference ones for entropy calculation.

The divergence entropy measuring the "distance" between two distributions can be calculated according to the following definition [10]:

$$D_{KL}(p|p^0) = \sum_{i=1}^{N} p_i \log_2 (p_i / p_i^0), \quad (4)$$

where: D_{KL} – distance entropy (Kullback-Leibler distance entropy), p – probability of a particular observed event (hydrophobicity density), p^0 – probability in the reference distribution. The index "i" denotes a particular amino acid. N denotes the number of amino acids in the polypeptide chain.

The analyzed distribution (hydrophobicity density transformed to probability density) (O) as it appears in the molecule under consideration is put as (p_i). The target distribution is the theoretical one (T) and to make possible interpretation of entropy values the random distribution (Rd). The random distribution is constructed using all p_i^0 values as equal 1/N, where N is the number of residues in the polypeptide chain. The protein characterized by O/T < O/Rd (observed versus theoretical in relation to observed versus random one) is assumed to represent the hydrophobic core structure accordant with the 3-D Gauss function. This means that its hydrophobic core is accordant with the idealized one.

Infrastructure

To perform the calculations we used the Grid infrastructure, provided by EUChinaGrid project. The EUChinaGRID pilot in-

frastructure at the time of the test run comprised 7 Grid sites, 5 from Europe and 2 from China, with a total amount of over 600 CPUs. We used gLite grid middleware, which was integrated with our portal and experiment management system [11]. Using our simulation and job management software we reached the average processing rate of ca. 300 structures per day. The total amount of produced data size was 150 MB of compressed PDB files. After the experiments, the simulation software has been integrated with the ViroLab virtual laboratory [12, 13].

Results and discussion

A comparative analysis of the R and S structures was presented in [8] An additional analysis consisting in a quantitative comparison between the R and S structures based on the divergence entropy [9] was applied and is presented in this paper. The divergence entropy allows comparison of two distributions in terms of distance between them. The hydrophobicity distribution in form of a profile along the polypeptide chain is the focus of the analysis.

Theoretical versus random distribution in R and S structural forms

The resultant structures for R and S forms were analyzed with respect to the hydrophobic core construction. The divergence entropy to express the distance between observed (O) and theoretical (T) hydrophobicity density distribution was calculated. Since the entropy values can be interpreted only in relative form a random distribution (Rd) has been taken under consideration. The proteins characterized by O/T < O/Rd were treated as representing the hydrophobic core structure accordant with the expected one. An opposite relation identifies the protein characterized by a tertiary structure not accordant with the expected 3-D Gauss function. Among 20 selected structures generated according to "fuzzy oil drop" model, 13 appeared to display a structure of the hydrophobic core accordant with the assumed model. The remaining 7 S structures not accordant with the model are those which generally failed to be folded using "fuzzy oil drop" model. Among 20 R structures generated according to Rosetta *abinitio*, 6 displayed a "fuzzy oil drop" like structure of the hydrophobic core, although no external force field was applied in these cases. The results are shown in Table 1.

	S		R	
	O/T	O/Rd	O/T	O/Rd
102	**0.087**	**0.288**	0.150	0.128
372	**0.134**	**0.344**	0.121	0.109
386	0.619	0.206	**0.114**	**0.153**
435	**0.069**	**0.229**	**0.059**	**0.145**
438	0.433	0.229	**0.166**	**0.199**
595	**0.096**	**0.274**	0.157	0.103
913	**0.143**	**0.195**	**0.098**	**0.147**
1000	**0.073**	**0.264**	0.113	0.084
1056	**0.093**	**0.218**	0.131	0.111
1134	0.790	0.164	0.143	0.129
1167	0.244	0.236	0.176	0.135
1281	**0.115**	**0.178**	0.145	0.133
1349	**0.082**	**0.244**	0.147	0.103
1356	**0.082**	**0.226**	0.114	0.098
1570	**0.177**	**0.221**	**0.119**	**0.128**
1736	**0.254**	**0.268**	0.145	0.103
2265	1.230	0.136	0.162	0.092
2300	**0.138**	**0.242**	**0.074**	**0.111**
2748	0.482	0.201	0.202	0.104
3208	0.572	0.156	0.185	0.118

Tab. 1. The divergence entropy measuring the distance between the hydrophobicity distribution as it appears in observed versus theoretical and observed versus random distribution. The structures characterized by the relation O/T < O/R are treated as structures accordant with the idealized distribution and distinguished in bold.

R to S comparison

The standard comparison to calculate the structural similarity expressed by RMS-D values was performed. The idealized (T) distribution generated for R and S structural forms as well observed (O) distributions were taken to measure the structural similarity in respect to the hydrophobic core construction. The final results are shown in Table 2.

The similarity of the hydrophobicity distribution in molecule 2300 (selected as the example of high accordance between R and S forms) can be seen in Fig. 1.

The large values of RMS-D appeared when the polypeptide chain was unfoldable using the "fuzzy oil drop" model [8]. An interesting observation is that those structures which failed to be folded according to the "fuzzy oil drop" model appeared to be folded in their R form and to be characterized by a 3-D Gauss-like hydrophobic core (structures 386, 438)

PROTEIN	RMS-D R-S	Correlation coefficient between $\Delta\tilde{H}$ values	THEORETICAL		OBSERVED	
			R/S	S/R	R/S	S/R
102	**7.562**	**0.365/0.286**	**0.319**	**0.344**	0.291	0.293
372	**12.139**	**0.510/0.461**	**0.291**	**0.332**	0.433	0.374
386	23.139	**0.408/0.372**	0.653	0.539	**0.173**	**0.157**
435	**10.140**	**0.487/-**	**0.220**	**0.243**	**0.263**	**0.247**
438	28.222	0.205/-	0.827	0.619	**0.163**	**0.147**
595	**9.687**	**0.453/0.453**	**0.239**	**0.223**	0.302	0.288
913	14.866	-0.133/-0.148	**0.418**	**0.395**	**0.220**	**0.209**
1000	**7.077**	**0.389/-**	**0.281**	**0.300**	0.330	0.294
1056	**7.848**	**0.343/-**	**0.311**	**0.321**	0.289	0.283
1134	25.990	0.148/-	0.951	0.697	0.113	0.100
1167	19.861	0.265/0.323	0.485	0.456	0.242	0.192
1281	**8.955**	**0.502/0.516**	**0.301**	**0.298**	0.182	0.176
1349	**7.493**	**0.432/0.411**	**0.182**	**0.181**	0.202	0.192
1356	**7.628**	**0.280/0.289**	**0.362**	**0.344**	0.214	0.193
1570	15.409	**0.101/0.161**	**0.503**	**0.472**	**0.177**	**0.177**
1736	20.011	0.027/-	**0.586**	**0.542**	0.301	0.279
2265	31.807	0.111/-	1.046	0.761	0.167	0.165
2300	**6.693**	**0.489/-**	**0.158**	**0.159**	**0.292**	**0.257**
2748	18.899	**0.254/-**	0.479	0.552	0.213	0.185
3208	23.033	0.158/0.159	0.787	0.646	0.153	0.132

Tab. 2. Divergence entropy measuring the differences between two distributions (hydrophobicity density distribution) for theoretical distribution as calculated for R and S version (left part of table) and for observed distribution as calculated for R and S structural forms. The ten structures characterized by lowest RMS-D values are given in bold.

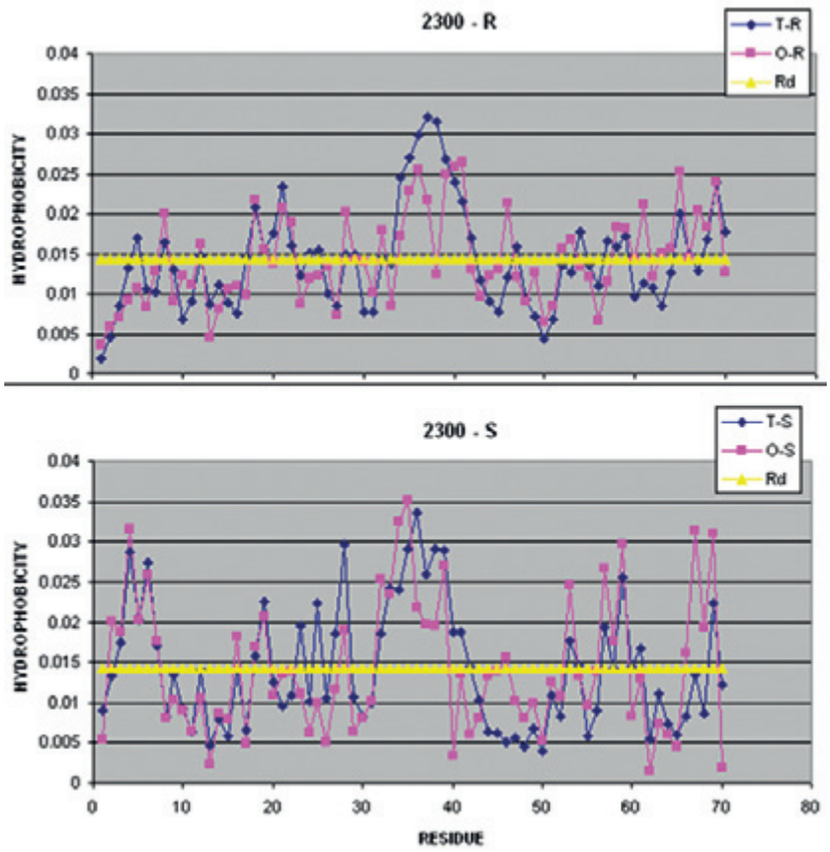


Fig. 1. Theoretical versus observed and random hydrophobicity distribution as measured in the R and S structural forms of protein 2300.

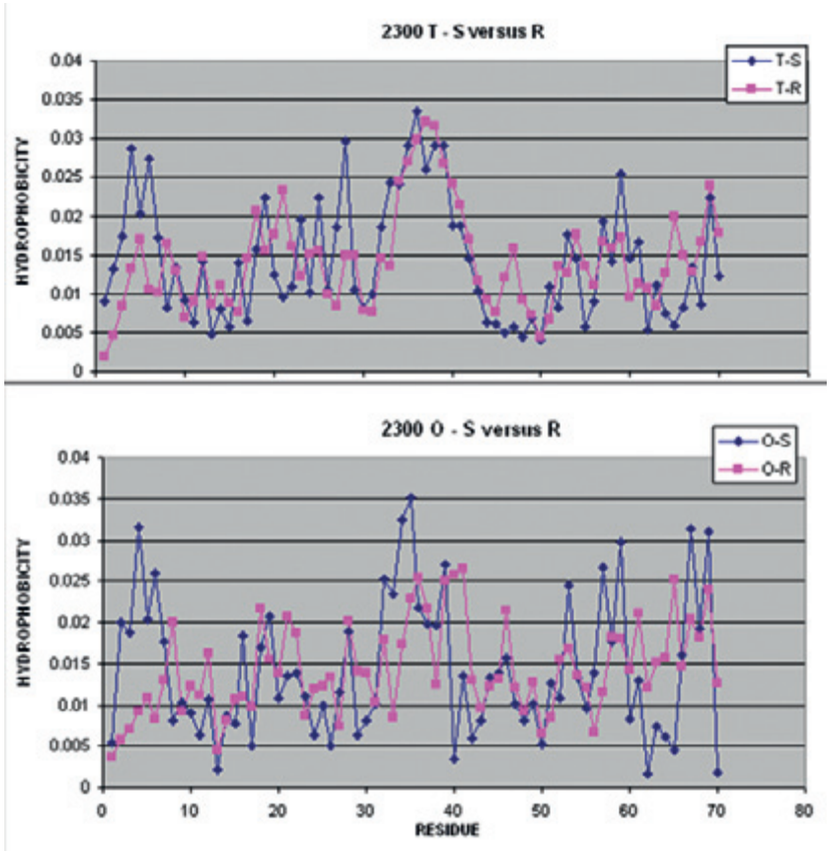


Fig. 2. Theoretical (T) (upper) and observed (O) (lower) hydrophobicity distribution in R and S structural forms.

The quite high similarity of hydrophobicity distribution can be seen in Fig. 1 The interesting observation is that both approaches identified the residues # 30 – 40 as the hydrophobic core. High similarity between theoretical and observed distribution can be seen especially in R form.

The degree of accordance between expected and observed hydrophobicity distribution can be seen in Fig.2. The expected distributions in both approaches seem to be highly accordant however the observed ones differ more. The interesting characteristic is the selection of the # 30-40 residues in both approaches as the fragment generating the hydrophobic core. Just the N- and C-terminal fragments seem to differ in both approaches. Differences in the observed distributions seem even to be higher than in expected ones.

An example of rather large differences between R and S form are shown in Fig. 3

An high accordance between the expected and observed distribution can be seen for the R form. The S form displays a much lower similarity between these two distributions. The hydrophobic core in the R form is generated by the residues 20-30 and 50-60 while in the S form the hydrophobic core is generated by the residues 40-60. Also the accordance between expected and observed hydrophobicity distribution is higher in the R form.

The distance entropy appeared to be strongly correlated with RMS-D values for appropriate structures of R and S structural forms. The theoretical (idealized) hydrophobicity distribution seems to represent the general feature of the protein molecule. The correlation between RMS-D and distance entropy for observed hydrophobicity distribution is much lower (regression function not significant). It can be explained as the effect of a large number of R structures recognized as structures close to random distribution.

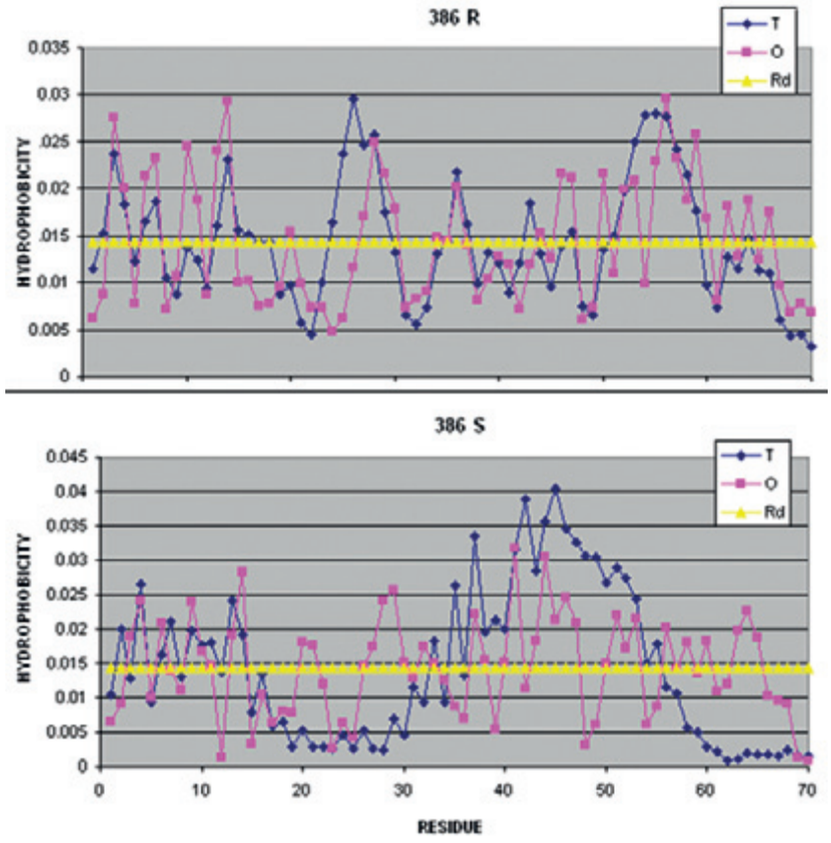


Fig. 3. Theoretical (T) versus observed (O) and random (Rd) hydrophobicity distribution as observed in the R (upper) and S (lower) forms of protein 386.

PROTEINS		FORM S	
		ACCORDANT	DISACCORDANT
FORM R	ACCORDANT	**435, 2300,** 913, 1570	386, 438
	DISACCORDANT	**102, 372, 595, 1000, 1056, 1281, 1349, 1356,** 1736	1134(U) , 2748,1167, 3208 (U) , 2265 (U)

Tab. 3. Summary of the "fuzzy oil drop" based comparison of structures generated as R and S. The numbers of structures given in bold represent the low RMS-D values. The symbol (U) denotes the structures which failed to be folded with the S approach.

Table 3 gives the summary of the results showing the structures accordant with "fuzzy oil drop" model in both R and S models. The low RMS-D structures appeared to be accordant with the "fuzzy oil drop" model. The sequences which failed to be folded according to the "fuzzy oil drop" model appeared to be represented by structures not accordant with the assumed model also for R structural forms.

The proteins folded according to the *ab initio* model as applied in ROSETTA appeared to represent the order accordant with the "fuzzy oil drop" model although in a much lower number (only 4 out of 20 structures). The unfoldable structures for obvious reasons do not display a hydrophobic core structure accordant with the 3-D Gauss function.

Statistical analysis

The relation between the RMS-D and entropy based similarity measurements are shown in Fig.4. The results of statistical calculations are given in Table 4.

A very good agreement of RMS-D and D_{KL} similarity values can be observed for theoretical hydrophobicity distribution. The high RMS-D values represent the structures which appeared unfoldable S structures. This is why significantly different is also the expected hydrophobicity distribution.

Low (statistically not significant using the regression function analysis) is the relation between entropy-based and RMS-D measurements for observed distributions. The low values of entropy for large values of RMS-D are the results of highly discordant observed (O) hydrophobicity versus theoretical (T) one.

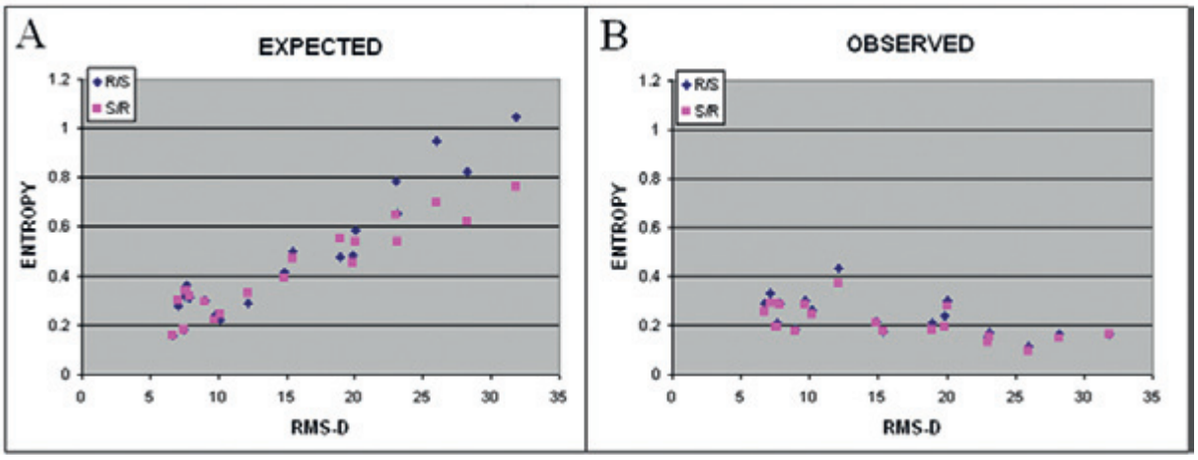


Fig. 4. The relation between two measurements of structural similarity: RMS-D and distance entropy based scale for A – theoretical (expected) (T) and B – observed (O) hydrophobicity distributions calculated for S and R structures.

COMPARISON	N	R – Spearman	t(N-2)	p
(RMS-D) R-S & (T) S/R	20	0.880782	7.89160	0.000000
(RMS-D) R-S & (T) R/S	20	0.893233	8.42909	0.000000
(RMS-D) R-S & (O) R/S	20	-0.628571	-3.42886	0.002994
(RMS-D) R-S & (O) S/R	20	-0.682211	-3.95864	0.000921

Tab. 4. Statistical analysis expressed by correlation coefficients measuring the relation between particular similarity measurements. All pair-wise comparisons appeared statistically significant.

The accordance of two similarity measurements (RMS-D and D_{KL}) appeared statistically significant. It means that the D_{KL} scale can be used for similarity search. The comparison of entropy values is possible in this case due to the equal number of amino acids in polypeptides (the entropy value depends on the length of the polypeptide).

The analysis of the data given in Figure 4 suggests that the theoretically expected structure of hydrophobic core is specific for a particular structure. The discordance between observed and theoretical hydrophobicity distribution results in a non significant relation between D_{KL} and RMS-D similarity measurements.

	Slope Intercept	Std. error	p-value
(T) S/R	-2.6786 42.73698	1.662360 3.655216	0.124602 0.000000
(T) R/S	1.62860 2914002	1.199315 2.244279	0.191262 0.000000
(O) R/S	29.6130 -60.5510	4.98566 20.14692	0.000013 0.007595
(O) S/R	31.0368 -72.4140	4.85207 21.35808	0.000005 0.003260

Tab. 5. The statistical analysis of the regression function expressing the relation between RMS-D and DKL structural similarity measurements.

Conclusions

In a previous study a set of proteins of 70 amino acids in polypeptide chain was selected from PDB. These proteins were classified according to their biological activity [14-16]. The accordance of their structure was analyzed in the context of the "fuzzy oil drop" model and it was shown that the proteins belonging to antifreeze proteins display a structure accordant with the idealized hydrophobicity distribution [16]. This is why the model was applied to the structural analysis of proteins whose structures were generated using *in silico* tools.

The "fuzzy oil drop" model applied for molecular dynamics simulation of a transmembrane protein appeared to describe the dynamic behavior of this protein in high agreement with the all-atoms simulation [17]. However the computing time required was significantly lower for simulation in the presence of an external force field of hydrophobic character expressed by a 3-D Gauss function.

The structures generated for NBPs using the two discussed models appeared highly similar with respect to molecule topology [18].

The presence of ordered hydrophobic core in proteins of structure generated *in silico* seems to support the reliability of the "fuzzy oil drop" model. Particularly for structures generated using ROSETTA with no initial assumption on the presence of an ordered hydrophobic core in protein structure.

The grid infrastructure seems to play the significant role for large scale computing in pharmacology [19].

Acknowledgements

This work has been supported by the European Commission EUChinaGRID project grant (contract number: 026634).

References

1. CASP, http://predictioncenter.org/ (Jan 10. 2011).
2. Chiarabelli C., Vrijbloed J.W., Thomas R.M., Luisi P.L.: Investigation of de novo totally random biosequences. Part I: A general method for in vitro selection of folded domains from a random polypeptide library displayed on phage. *Chem Biodiversity* 2006, 3: 827-839.
3. Chiarabelli C., Vrijbloed J.W., Lucrezia D.D., Thomas R.M., Stano P., Polticelli F., Ottone T., Papa E., Luisi P.L.: Investigation of de novo totally random biosequences. Part II: On the folding frequency in a totally random library of de novo proteins obtained by phage display. *Chem Biodiversity* 2006, 3: 840-859.
4. Rohl C.A., Strauss C.E.M., Misura K.M.S., Baker D.: Protein structure prediction using Rosetta. *Methods Enzymol* 2004, 383: 66-93, http://boinc.bakerlab.org/rosetta/.
5. Evangelista G., Minervini G., Luisi P.L., Polticelli F.: RandomBLAST a tool to generate random "never born protein" sequences. *Bio-Algorithms and Med-systems* 2007, 3: 27-31.
6. Minervini G., Evangelista G., Villanova L., Slanzi D., De Lucrezia D., Poli I., Luisi P.L., Polticelli F.: Massive non-natural proteins structure prediction using grid technologies. *BMC Bioinformatics* 2009, 10, Suppl. 6: 22.
7. Konieczny L., Brylinski M., Roterman I.: Gauss-function-based model of hydrophobicity density in proteins. *In Silico Biology* 2006, 6: 15-22.
8. Prymula K., Piwowar M., Kochanczyk M., Flis L., Malawski M., Szepieniec T., Evangelista G., Minervini G., Polticelli F., Wisniowski Z., Salapa K., Matczynska E., Roterman I.: In silico structural study of random amino acid sequence proteins not present in Nature. *Chemistry & Biodiversity* 2009, 6: 2311-2336.
9. Levitt M.: A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol*. 1976, 104: 59-107.
10. Nalewajski R.: Information theory of molecular systems, Amsterdam [etc.]: Elsevier, 2006.
11. Malawski M., Szepieniec T., Kochanczyk M., Piwowar M., Roterman I.: An approach to protein folding on the grid – EUChinaGrid experience. *Bio-Algorithms Med-Systems* 2007, 3, 5: 45-49.
12. Bubak M. et al.: Virtual Laboratory for Collaborative Applications. In: M. Cannataro (ed.), Handbook of Research on Computational Grid Technologies for Life Sciences. Biomedicine and Healthcare. Chapter 27, pp. 531-551. Information Science Reference, 2009, IGI Global.
13. Bubak M., Malawski M., Rycerz K., Turala M.: CrossGrid – Tools and Services for Interactive Grid Applications. *TASK QUARTERLY*. Scientific Bulletin of Academic Centre in Gdansk. TASK Publishing, 2004, 8: 503-512.
14. Prymula K., Roterman I.: Functional characteristics of small proteins (70 aa) forming protein-nucleic acid complexes. *J. Biomol. Struct. Dynam*. 2009, 26: 663-677.
15. Prymula K., Roterman I.: Structural entropy to characterize small proteins (70 aa) and their interactions. *Entropy* 2009, 11: 62-84.
16. Prymula K., Salapa K., Roterman I.: "Fuzzy oil drop" model applied to individual small proteins build of 70 amino acids. *J. Mol. Model.* 2010, 16: 1269-1282.
17. Zobnina V., Roterman I.: Application of the fuzzy oil drop model to membrane protein simulation. *Proteins* 2009, 77: 378-394.
18. Minervini G., Evangelista G., Polticelli F., Piwowar M., Kochanczyk M., Flis L., Szepieniec T., Wisniowski Z., Matczynska E., Prymula K., Roterman I.: Never born proteins as a test case for ab initio protein structures prediction. *Bioinformation* 2008, 3: 177-179.
19. Roterman I.: E-infrastructure technologies triggering of bioinformatics development. *Bioinformation* 2007, 2: 126-127.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 13-16

SIMPLIFYING ASSUMPTIONS AND THE SCOPE OF APPLICATION OF LIPID MEMBRANE MODELS

KRYSTIAN KUBICA

Department of Biomedical Engineering and Instrumentation, Wrocław University of Technology, 50-370 Wrocław, Poland

Abstract: Biological membranes are components of the cell – the basic unit of life. Their structure originates from amphiphilic properties of lipid molecules (the major constituent of biological membranes) which when surrounded by water spontaneously form organized structures that include bilayers. A bilayer created purely out of lipid molecules is used as a physical model of biological membranes on which one can study biological processes associated with their lipid phase. These may include: passive transport, formation and decomposition of domains, phase transitions and formation of pores influenced by an external electric field. Experiments on lipid bilayer coupled with studies of its mathematical models enable to gain an understanding of the aforementioned biological membrane phenomena at a molecular level. A mathematical model is characterized by a set of simplifying assumptions which determines its application. By developing a simple model that only takes into account the structure of the hydrophobic lipid molecules, we were able to observe phases of various density corresponding to temperature changes. Expanding the model by including the polar parts of lipid molecules expressed via a surface pressure multiplied by a surface area per one molecule increased its range of research. This assumption, did not allow capturing some of the factors such as ionic strength or a presence of water molecules. Supplementing the model with new assumptions increased its application. The extended model allowed additionally tracking changes in the membrane influenced by biologically active amphiphilic compounds as well as examining the process of electroporation.
Keywords: model of biological membrane, membrane properties, assumption of model

Introduction

Biological membranes are important constituents of the cell. The structure of biological membranes is a consequence of the amphiphilicity of lipid molecules, which form organized structures such as bilayers in an aqueous environment. Due to biological membranes being complex, it is more straightforward for some of the studies to being conducted on physical models (lipid bilayers). Current techniques allow studying lipid bilayers in the form of liposomes, macrovesicles or flat membranes. Theoretical studies of mathematical models allow supplementing a description of molecular mechanisms and, conspicuously, as expected from a mathematical modeling, make predictions possible. Each model is based on simplifying assumptions that delimit a scope of its application.

Experimental verification of the hypotheses that arise from the analysis of theoretical models allows accepting a given model or restricting the scope of its application, and even rejecting it. A fast development of computational techniques makes them a perfect tool for theoretical studies of the models. The computer simulation techniques, based on molecular dynamics, make it possible to track an evolution of a system of hundreds of molecules during some tens of nanoseconds. Another type of a theoretical analysis of the complicated lipid systems is based on statistical models.

Theoretical models of lipid membranes

Statistical models have been allowing studying, yet not without certain limitations, membrane processes since 1970s. Scott [1] simplified the structure of lipid molecules (Fig. 1) to study phase transitions between the lipid phases termed *kink* and *gauche* blocks. He showed that such transitions may occur when only taking into account van der Waals interactions and a conformational energy of lipid chains (neglecting an interaction energy of the polar parts).

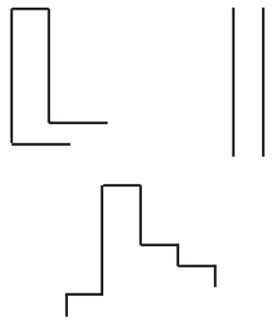


Fig. 1. The three out of the twenty-one assumed lipid conformers that build *gauche* and *kink* blocks [1]. The three C-C bonds in *gauche, trans* and *gauche* conformations were called a *kink* (marked as a step).

Based on Scott's model an optimal length of alkyl chain in a homologous series of intruder molecules (4, 6, 8, 10, 12 and 14 carbon atoms in a chain) that modifies membranes composed of lipids with alkyl chains of only 14, 16 or 18 carbon atoms was found. The maximum amount of relative *gauche* blocks was observed for membranes modified by intruder molecules with the optimal chain length [2]. Similar properties were captured experimentally when studying permeability of liposomes modified with the amphiphilic compounds having 10,12,14 and 16 carbon atoms, and belonging to the homologous series of N-alkoxymethylene-N,N,N-trimethyl-ammonium chlorides and N,N,N-trimethylglycine n-alkylesters [3]. This observed qualitative conformity between the theoretical and physical models allow thinking that the presence of optimum alkyl chain length in the homologous series of modifier molecule is mainly due to van der Waals interactions between the chains. The interaction of the polar parts may affect the depth of modifier molecule incorporation into the lipid membrane.

The Pink's model published in 1980s [4, 5] included the polar parts of lipid molecules via the product of a surface pressure and an area per one molecule i.e. per one alkyl chain. This model has become the basis for many studies utilizing Monte Carlo simulation methods. According to Pink's model, the membrane is modeled with a triangular net which nods have six equivalent close neighbors. A lipid molecule composed of two chains occupied two neighboring nods. Each chain occupied one out of ten defined conformations, and was treated independently. Such assumption did not tie specific two chains in one lipid molecule. Individual chains were characterized by its conformation states (always 1 out of 10 possible). Each state was defined by a conformational energy, a degree of degeneration and an area per chain. A total energy of a lipid layer was determined by the Hamiltonian:

$$H = -\frac{J_o^M}{2} \sum_{<i,j>} \sum_{nm} I_{nm}(r_{nm}) L_{ni} L_{mj} + \sum_i \sum_n (\pi A_n + E_n) L_{ni} \quad (1)$$

where:

J_o^M – a van der Waals interaction constant,
L_{ni} – a lipid chain projection operator for lattice,
$<i,j>$ – a sum over nearest-neighbor sites,
p – an internal pressure,
E_n – an internal energy of a chain in state n
r_{nm} – a distance between two chains in states n and m at sites i and j
A_n –an area per chain,

Pink's model required setting values of two parameters: J_o^M and p which describe van der Waals energy of two neighboring chains being in the all-trans conformation and the internal pressure. They were selected to reflect the experimental gel-fluid phase transition temperature. The detailed analysis of the assumptions was given earlier [6]. Based on Pink's model many membrane aspects have been studied: gel-fluid phase transition [7], lipid membrane organization [8], the influence of biological active compounds on membrane properties [9, 10]. Studies on the spatial structure of lipid molecules within membranes [11] have shown that the two initial carbon atoms of β-chain of the phosphatidylcholine molecule lay in the membrane plain. Assuming that one alkyl chain is effectively shorter, and assigning one lipid molecule to two neighboring nods in the hexagonal net, it was observed that the gel-fluid phase transition temperature decreases by 10°C. The value of the surface pressure parameter p cannot be constant for the membranes formed out of lipids with various chain lengths [6]. A 4% change in the value of parameter π caused a change in the gel-fluid phase transition temperature by 1°C. Pink's model did not take into account the polar environment of the model membranes such as ionic force, pH or number of water molecules. The dielectric constant changes from 80 for the polar part of membrane to 2 for the hydrophobic layer [12]. When we assume that the lipid polar head has a dipole structure an importance of the mentioned environment parameters may be studied [13]. A further enrichment of Pink's model with new assumptions enhanced the research scope of the model. According to the assumptions the polar head was able to rotate and change its slope in relation to the membrane surface. Furthermore, each lipid molecule was able to shift along the normal to the plane (Fig. 2). These assumptions enabled modeling the gel-fluid phase transition and in addition the undulated lipid phase [13].

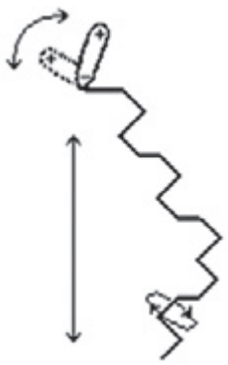


Fig. 2. According to the model assumption, one lipid molecule has one chain and one head – a dipole that can rotate and change its tilt. The whole molecule can be shifted perpendicular do the membrane plane.

Unfortunately, the defined model was not susceptible to changes of ionic strength of the medium at temperatures higher than that of the main phase transition, which was confirmed ex-

perimentally [14]. A fluorescence probe (phosphatidyletanolamine with a covalently bound fluoresceine – fluoro-PE), incorporated into a series of liposomes formed out of DMPC, DPPC and DSPC lipids displayed changes in its fluorescence properties that depended on the ionic strength of the medium only at temperatures exceeding the main transition temperature. Experimental results forced a modification of the theoretical model. A detailed analysis of the dipole system with the assumption letting to imitate the reality better, i.e. that one dipole belongs to two hydrocarbon chains (Fig. 3) [15, 16], showed that the qualitative behavior of the system changes when the dipole electric charge is less than 0.5 of the elementary charge. Such restriction showed that at temperatures below the gel-fluid phase transition temperature the dipole system did not exhibit the significant changes with an ionic force variation (for 10mM and 100mM). At temperatures higher than the gel-fluid phase transition temperature, a 10-fold decrease in the ionic force caused an increase in the number of vertical dipoles. At 300 K only 14% of dipoles took the vertical position and at 330K the number of vertical dipoles increased to 50%.

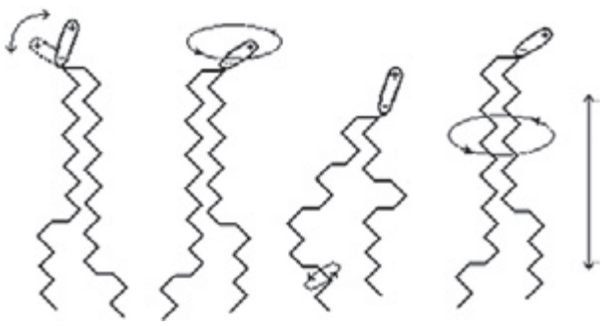


Fig. 3. Assumed degrees of freedom of lipid molecules describe head tilt, head rotation, rotation around -C-C- bound, rotation and shifting of the whole molecule.

A fluorescence intensity of the probe depends on a local pH value which can be expressed by pH of the further neighborhood and a local value of the electric potential [17]. The local electric potential depends on configuration of the polar parts of lipid molecules – dipoles. Analysis of the interaction energy between the probe molecules and the lipid molecules of the membrane showed that under certain experimental conditions single probe molecules can enforce the changes in the configuration of neighboring lipids, affecting the local pH value and the fluorescence intensity [18].

Pink's model extended with the new assumptions gives the possibility to study the lipid systems under the electric field. Since the beginning of the last century effects of the electric field on the properties of cell membranes has been known. Experimental investigations have shown that the electric field of adequate parameters can cause membrane electroporation [19, 20]. This phenomenon, which a molecular mechanism is not yet precisely known, finds practical applications. Electroporation, among others, is used as a factor facilitating drug entry into the cell. Studies of the theoretical model can predict molecular changes that result in a pore formation [21, 22]. Computer simulation performed on the latter model subjected to the action of the electric field gave threshold value of the electric field strength of 10^6V. Higher electric fields forced a vertical orientation of the polar heads and relaxation of the hydrocarbon chain packing in the lipid layer on the negative side of the field. The same electric field forced a lying position of the polar heads and increased the chain packing in lipid layer on the positive side. No changes were observed for electric fields smaller than the threshold value. The differences in packing density of both layers could be removed by a transfer of only one chain from lower to upper layer for every 50 chains. This process caused also a change in the position of the lipid's dipole – it was located in the hydrophobic part of membrane. The dipole surrounded by water molecules facilitated the formation of a hydrophilic pore across the membrane. The dependence of lipid molecules packing density in a bilayer on spatial configuration of their polar parts (dipoles) is consistent with earlier studies of the theoretical model [15].

References

1. Scott H.L.: Some models for lipid bilayer and biomembrane phase transitions. *J. Chem. Phys.* 1975, 62.
2. Kubica K., Kuczera J.: The effect of the alkyl chain length of modifier molecules on membrane structure studied by the cellular-automaton method. *Appl. Math. Comp.* 1991, 44.
3. Kuczera J., Janas T., Przestalski S., Witek S., Oświęcimska M.: Effect of surface active anions on the activity of dodecyloxymethylenemethylmorpho-linium chloride as modifier of sulphate ion transport across lecithin liposome membranes. *Studia Biophysica* 1985, 105.
4. Caille A., Pink D., De Verteuil F., Zuckerman M.: Theoretical models for quasi-two-dimensional mesomorphic monolayers and membrane bilayers. *J. Can. J. Phys.* 1980, 58.
5. Pink D., Green T., Chapmann D.: Raman Scattering in Bilayers of saturated phosphatidylcholines. Experimental and theory. *Biochemistry* 1980, 19.
6. Kubica K.: Pink's model and lipid membranes. *Cell. Mol. Biol. Lett.* 1997, 2.
7. Mouritsen O.G., Boothroyd A., Harris R., Jan N., Lookman T., MacDonald L., Pink D.A.: Computer simulation of the main gel-fluid phase transition of lipid bilayers. *J. Chem. Phys.* 1983, 98.
8. Ipsen J., Mouristen O., Bloom M.: Relationships between lipid membrane area, hydrophobic thickness, and acyl-chain orientational order. *Biophys. J.* 1990, 57.
9. Jorgensen K., Ipsen J., Mouritsen O., Bennett D., Zuckermann M.: The effect of density fluctuations on the partitioning of foreign molecules into lipid bilayers: application to anaesthetics and insecticides. *Biochim. Biophys. Acta* 1991, 1067.
10. Sujak A., Okulski W., Gruszecki W.I.: Organisation of Xanthophyll Pigments Lutein and Zeaxanthin in Lipid Membranes Formed with Dipalmitoylphosphatidylcholine. *Biochim. Biophys. Acta* 2000, 1509.
11. Hauser H., Pascher I., Sundell S.: Conformation of phospholipids crystal structure of a lysophosphatidylcholine analogue. *J. Mol. Biol.* 1980, 137.
12. Brockman H.: Dipole potential of lipid membranes. *Chem. Phys. Lipids* 1994, 73.

13. Kubica K.: Monte Carlo simulation towards ripple phase modeling. *Comp. Chem.* 2001, 25.
14. Langner M., Pruchnik H., Kubica K.: The effect of the lipid bilayer state on fluorescence intensity of Fluorescein-PE in a saturated lipid bilayer. *Z. Naturforsch.* 2000, 55c.
15. Kubica K.: Computer simulation studies on significance of lipid polar head orientation. *Comp. Chem.* 2002, 26.
16. Kubica K.: Computer simulation studies on the significance of lipid polar head charge. *Cel. Mol. Biol. Lett.* 2002, 7.
17. Westman J., Boulanger Y., Ehrenberg A., Smith I.C.P.: Charge and pH dependent drug binding to model membranes. A H-NMR and light absorption study. *Biochim. Biophys. Acta* 1982, 685.
18. Kubica K., Langner M., Gabrielska J.: The dependence of Fluorescein-PE fluorescence intensity on lipid bilayer state. Evaluating the interaction between the probe and lipid molecules. *Cel. Mol. Biol. Lett.* 2003, 8.
19. Weaver J., Chizmadzhev Yu. A.: Theory of electroporation: A review. *Bioelectrochem. Bioenerg.* 1996, 41.
20. Kalinowski S., Koronkiewicz S., Kotulska M., Kubica K.: Simulation of electroporated cell by chronopotentiometry. *Bioelectrochemistry* 2007, 70.
21. Kotulska M., Kubica K.: Structural and energetic model of the mechanisms for reduced self-diffusion in lipid bilayer with increasing ionic strength. Physical Review E 72, 2005.
22. Kubica K.: A pore creation in a triangular network model membrane. *Comp. Biol. Chem.* 2008, 32.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 17-23

APPLICATION OF NUMERICAL ANALYSIS OF FLUORESCENCE SPECTRUM TO IDENTIFY PROPERTIES OF SUBSTANCES ASSOCIATING WITH CONGO RED MICELLE

Grzegorz Zemanek, Anna Jagusiak, Dorota Kusior, Barbara Piekarska, Paweł Spólnik, Barbara Stopa

Chair of Medical Biochemistry, Medical College, Jagiellonian University, Kopernika 7, 31-034 Krakow, Poland

Abstract: The particles of Congo red (CR), a bis-azo dye, associate in aqueous solutions, forming ribbon-like micelles. Supramolecular dyes, as a result of their molecular size, have a limited capability of penetrating into the interior of the majority of native proteins and do not enter living cells surrounded by an intact phospholipid membrane. They may however, bind to proteins with structure destabilized by unfolding (during denaturation) or function-related changes (e.g. surface cell receptors). Congo red absorbs visible light with a maximum absorption at 500 nm, in aqueous solutions at neutral pH. After CR bind to the polymers (such as cellulose) or certain proteins, dye's behaviour changes remarkably. These binding affects spectral properties of the dye: the absorption spectrum of the Congo red shifts towards longer wavelengths and receives a yellow-red fluorescence. A numerical analysis of the graphic data obtained from fluorescent microscope (division into channels corresponding to constituent colours and with the noise separated using a graphical program) allowed to understand better the causes of movement of the CR fluorescence spectrum and to identify the properties of substances associating with the CR micelle. By means of a statistical analysis of the data it was possible to observe that wavelength of emission and intensity of fluorescence is not only associated with the polarity of the environment but mainly with the strength of the association between the Congo red micelle and the matrix.
Keywords: Congo red, fluorescence, numerical analysis, proteins, self-assembling dyes

Abbreviations:
- CR – Congo red
- HCV 29T – epithelial human bladder transformed cells
- J774A.1 – murine monocyte cell line
- U937 – human leukemic monocyte lymphoma cell line
- SRBC – sheep red blood cells
- anti-SRBC – antibodies against sheep red blood cells
- IgG – Immunoglobulin G

Introduction

Properties of the Congo red

Molecules of Congo red (CR), a bis-azo dye, in water solution associate with each other in an organized way to produce micelles. In these ribbon-like structures a charged amino- and sulfo- groups are exposed to hydrophilic solvent and capable of making hydrogen bonds. Hydrophobic core of a single molecule, consisting of aromatic rings, binds to aromatic rings of other dye molecules, mainly through π-π bonds interactions (1, 2). Additionally, the stability of the micelle can be improved by inorganic ions present in solution which screen charges on CR side groups. As a result of these interactions the molecules form a supramolecular structure (although liquid crystal consists of multiple particles linked to each other by noncovalent bonds, it demonstrates properties of a single, large molecule) (1, 2). Ribbon-like micelles, unlike spherical micelles, do not protect all fragments of hydrophobic molecules forming the supramolecular structure. Hydrophobic regions in Congo red micelle, exhibited towards the polar environment, can be identified. These regions enable CR to associate with exposed hydrophobic structures of polymers (e.g. proteins) (3). Shape and plasticity of the ribbon-like micelle, resulting from the lack of covalent bonds between monomer molecules, allows for an adjustment of a micelle to the specific binding place in the protein molecule. Consequently, association between the protein (polymer) and the ribbon micelle is particularly strong. Taking into consideration the specific micellar shape, it seems that the essential conditions for binding liquid crystal ligands to proteins is the presence of the beta sheets in

a polypeptide chain (4, 5). However, these conditions are not sufficient for most of the native proteins. Additional requirement is an earlier instability of the beta plate or the entire tertiary structure of a polypeptide chain (4-6). This instability may be caused by the partial unfolding of the protein molecule (1, 7) or its temporary function-related conformational changes (1, 8-11).

On the other hand, supramolecular dyes have limited ability to penetrate a hydrophobic core of native proteins soluble in polar solutions, in contrast to non-aggregating molecules of similar chemical structures. Furthermore, the presence of the polar side groups hinders the ability of Congo red micelles to penetrate lipid bilayer, composed of amphipathic molecules. Consequently, CR molecules very slowly penetrate living cells surrounded by intact phospholipid membranes (12). Owing to these properties, bis-azo dye supramolecules may be applied to study and visualize many processes taking place on the cell surface. Studies of Congo red influence on the behaviour of the surface cell receptors revealed that CR binds cell surfaces in places of cell junction, during association of monocytes and HCV 29T line cancer cells (12). This interaction was visualized by dint of Congo red fluorescence properties, after its association with the proteins.

General applications of Congo red (CR)

Azo dyes, with more than 2000 different substances being currently used, make up the largest and most versatile class of dyes. More than $8.0 \cdot 10^8$ kg of dyes are annually produced worldwide, of which 60–70% are azo dyes (13, 14). Their chemical structure is characterized by one or more azo groups (–N=N–) and they can be used to colour a large number of different substrates, such as textiles, papers, leathers, gasoline, additives, food and cosmetics (13). CR is an example of a widely used bis-azo dye. It is commonly used in textile industry to give wool and silk red colour with yellow fluorescence (15). In respect of CR binding to ß-D-glucans, such as cellulose and chitin, it can be used as a fluorescent and a bright-field dye (16). Due to such properties, CR can also be used to detect fungus cells. Ammoniacal Congo red has been used to distinguish between species of parasitic fungi, its fluorescence have also been used to reveal fungal infections in human tissue samples and cultures (16).

CR is a dye used in histology and histopathology to detect the amyloid proteins. The dye was introduced more than 80 years ago and its gold-green birefringence under polarized light has been the gold standard for amyloid detection ever since (17-20). Specific properties observed after Congo red association with amyloid fibrous aggregates are most frequently explained by the presence of the cross-β-structure (21, 22). Furthermore, spectral variations observed for CR binding to amyloid-like proteins are similar to properties of the dye dissolved in organic solvents. This observation leads to the suggestion that the dye molecules must be exposed to similar changes in the environment in both cases (23). Currently, a large number of amyloid ligands have been synthesized and most of these agents are derivatives of the Congo red. As shown by Klunk and co-workers, the key structural feature of CR might be the two acidic functional groups and the spacing between them (17, 20). Chrysamine G, a lipophilic analogue of CR, which also has two acidic functional groups with the same spacing between them as observed in Congo red, has also shown high binding affinity to protein aggregates (17, 19).

Fluorescence of Congo red

The luminescence for analytical purposes most useful is evoked by excitation of light, i.e. photoluminescence. Photoluminescence can be observed directly while exposing the preparation to the appropriate wavelength of light, as in fluorescence microscopy or while measuring electric current created by emitted light, as in spectrofluorimeter. An excited electron can come back to its original (ground) state by emitting a photon of light in three ways: by fluorescence, by phosphorescence and by delayed fluorescence (24). Congo red associated with proteins demonstrates red fluorescence, when excited by visual light (maximum excitation 498 nm in aqueous solution). This emission of light, characterized by high intensity and persistence, can be observed, among others, after the dye binding to amyloids (25). This property of CR was used to detect amyloid proteins. Furthermore, it was pointed out that the sensitivity of this method is much better than the classic method, where polarized light was used (26). Additionally, both methods can be used together, because the way of sample preparation is the same but only changes the way of detection. However, both detection methods can be performed using the same measuring device (e.g. a properly equipped microscope). Such a double test has yet another advantage: it combines a method of higher sensitivity (fluorescence measurement) and a method of greater specificity (polarized light microscopy measurement) (26). Unfortunately, Congo red binds also non-specifically to proteins; its affinity does not depend only on the protein structure. It clearly binds stronger to lysozyme (with α- helices and β-sheets) and albumin (with α- helices only) than to ribonuclease A (with α -helices, β-sheets and random coil) or insulin (with short α- helices), where it practically does not interact (13, 27). Studies indicate that both CR and Acid Red 2 (an azo dye C.I. Acid Red 2), which has a similar structure to CR, bind to plasma albumin mainly by hydrophobic interactions and hydrogen bonds. The fluorescence spectrum (ultraviolet light) of plasma albumin changes after association with these dyes (13, 27).

Identification of the particular conditions enabling Congo red molecules to emit visible light, may help to better understand the mechanism responsible for the dye association with both protein and non-protein ligands. It seems that the phenomenon of CR fluorescence may be caused by changes in the polarity of the molecular surrounding between polar solution and the hydrophobic core of a protein as well as the change in the arrangement of dye molecules associating with the ligand.

Materials and methods

Reagents

Congo red (Aldrich) and other chemical reagents used were of analytical grade. Solutions used for fluorescence studies were prepared by dissolution of Congo red in analytically pure water, methanol, ethanol and propanol. Final concentration of CR was

0,1mg/ml (for propanol saturated solution). Polyclonal anti-SRBC IgG were isolated from sera of rabbits immunised with sheep red blood cells (SRBC).

Association of Congo red with proteins and with cellulose

Cellulose powder and cellulose fibres were stained for 10 minutes at room temperature in aqueous solution of Congo red, of a final concentration 5 mg/ml. After staining, powder and fibres were washed three times in the physiological salt. Silk and cat's hair were initially heated for 10 minutes at 100°C with the reduction of disulphide bonds by 17% β-mercaptoethanol and then they were washed in the physiological salt. Both proteins were stained for 10 minutes with the supramolecular form of 10 mg/ml solution of Congo red at 100°C. After staining, aggregates were washed three times in the physiological salt.

Heat-aggregated IgG were obtained by heating saline solution of human IgG (15 mg/ml) at 63°C. IgG aggregates were stained by the supramolecular dye Congo red of a final concentration 10 mg/ml at 63°C. The excess of the dye was removed by filtration on Bio-Gel® P-10 column (saline solution). The solution was then centrifuged at 60 g for 10 minutes and the supernatant was used for further studies.

Cell lines

Epithelial human bladder transformed cells (HCV29T, Polish Cellular and Molecular Biology Network, UNESCO/PAN), murine monocytes cell line (J774A.1, ATCC TIB67) and human leukemic monocyte lymphoma cell line (U937) were cultured by biweekly passages in RPMI1640 medium with 5% FCS and regularly tested for *Mycoplasma* contamination.

Isolation of monocytes

Peripheral human blood mononuclear cells were isolated from EDTA-treated blood by standard Ficoll Isopaque (Pharmacia Uppsala) gradient centrifugation. Monocytes were then separated by counter-flow-centrifugal elutriation using a JE-5.0 system equipped with a Sanderson separation chamber (Beckman) (by courtesy of Prof. J. Marcinkiewicz, Chair of Immunology, Medical College, Jagiellonian University). Isolated monocytes of 85-90% purity with no selection of subsets [34, 38] were suspended in RPMI-1640 medium (Biochrom Berlin) with antibiotics. Suspensions of $I \cdot IO^6$ cells/ml in 2 mM glutamine, 10% FCS solution, were used for culturing (at 5% CO_2, 37°C).

Cell cultures

J774A.1 murine monocytes and HCV29T cell suspensions in RPMI-1640 medium were cultured in culture dishes (HCV29T cell at a cover slip at the bottom of dishes) at 5% CO_2, 37°C. After an initial incubation of culture dishes with HCV29T cell, 80% of the supernatant medium was removed and replaced with medium, containing peripheral human blood mononuclear cells (10^6 cells/ml) or J774A.1 murine monocytes (10^6 cells/ml). Then they were cultured further for 2, 4 and/or 6 hours (at 5% CO_2, 37°C).

SRBC (1 ml, $6.6 \cdot 10^8$ SRBC/ml) with 50µl anti-SRBC antibodies (1mg/ml) were added to the culture dishes with 5ml J774A.1 murine monocytes ($I \cdot 10^6$ cells/ml) and then the culture was continued for 2 hours.

Cell stained

In cell models, Congo red was added to the culture at final concentration 0.05 mg/ml, at the beginning of activation (by addition J774A.1 murine monocytes, peripheral human blood mononuclear cells or SRBC – anti-SRBC immune complex). Before observation, culture dishes (for HCV29T cell the cover slip with the cell) were gently rinsed twice with saline.

U937 cells were stained by heating in saline solution of CR (5 mg/ml) at 55°C and 70°C. The excess of dye was removed by centrifugation.

Microscopic studies

The distribution of yellowish-red fluorescence derived from Congo red immobilized by binding to proteins was examined in a Carl Zeiss Axioplan II fluorescence microscope with integrated camera Canon PC1200 and AxioVision 4 V 4.6.3.0. software. Further transformations were obtained by the analysis of images, using AdobePhotoshop, SigmaGel and Statistica 7 programs.

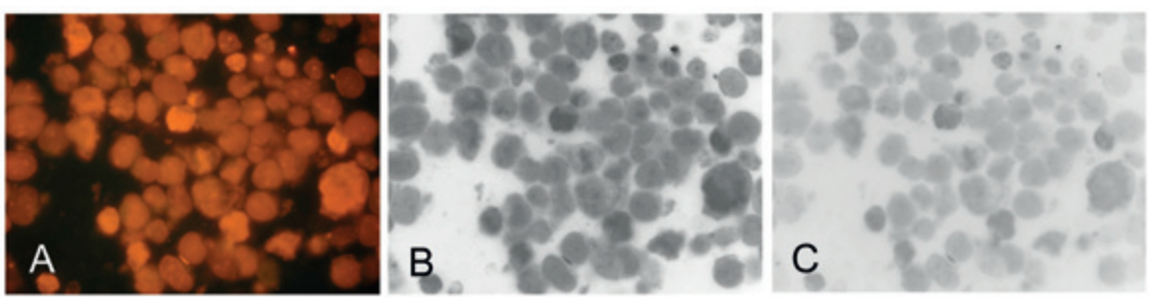


Fig. 1. Sample image analysis (cell line U937 with Congo Red): A – image from fluorescence microscope; B – separated fraction of red fluorescence after inversion; C – separated fraction of yellow fluorescence after inversion.

Measurement of solutions fluorescence

Congo red's fluorescence was measured by spectrofluorimeter Hitachi F-2000.

Results

In this paper the images of the following samples stained with Congo red were obtained: cellulose fibres and powder, heat-denatured aggregated immunoglobulin (IgG), silk, cat's hair, cell structures (e.g. epithelial cells HCV29T with human and murine J774A.1 monocytes, cell line J774A.1 with immunological complex SRBC-IgG anti SRBC) and the surface-denatured cells (cell line U937)**.** Images were observed using the fluorescence microscopy with excitation wavelength λ= 470 nm and then recorded with a digital camera. The images were separated into RGB colour channels (using a system separating colour images into three basic colours: red, green and blue) and converted to monochrome images using specialized algorithms (prepared on the basis of the monochrome masters). The separate monochrome images were generated for red and yellow colour and their intensity was changed by inversion (Figure 1).

The intensity of dark spots in the inverted pictures, which corresponds to the intensity of fluorescence, was measured with SigmaGel progam. After mathematical compensation of the background interference, the values of the fluorescence intensity were obtained for individual images and the fluorescence quotient (the ratio of red intensity to yellow intensity) was calculated. The

Comparison of fluorescence intensity quotients - red fraction's to yellow's - for the Congo red bound to polymers (protein, polysaccharide) and cells

The **Wilcoxon signed-rank test.** Results not marked in bold demonstrate statistical differences, p≥ 0.05

	Fibre cellulose	Cellulose powder	HCV/ human monocyte interaction	HCV/ mouse monocyte interaction	Monocytic cell line/IgG/RBC	U937 line after heat denaturation at 70°C	U937 line after heat denaturation at 55°C	Silk after heat denaturation at 100°C	Cat's hair after heat denaturation at 100°C	IgG after heat denaturation at 63°C
Fibre cellulose		0,0022	**0,4432**	**0,8613**	**0,3270**	0,0000	0,0000	0,0000	0,0000	0,0000
Cellulose powder	0,0022		0,0012	0,0277	0,0357	0,0000	0,0000	0,0000	0,0000	0,0000
HCV/ human monocyte interaction	**0,4432**	0,0012		**0,8613**	**0,6744**	0,0000	0,0000	0,0000	0,0000	0,0000
HCV/ mouse monocyte interaction	**0,8613**	0,0277	**0,8613**		**0,4008**	0,0015	0,0015	0,0015	0,0015	0,0015
Monocytic cell line/IgG/RBC	**0,3270**	0,0357	**0,6744**	**0,4008**		0,0117	0,0117	0,0117	0,0117	0,0117
U937 line after heat denaturation at 70°C	0,0000	0,0000	0,0000	0,0015	0,0117		0,0000	0,0000	0,0000	0,0000
U937 line after heat denaturation at 55°C	0,0000	0,0000	0,0000	0,0015	0,0117	0,0000		0,0000	0,0000	0,0000
Silk after heat denaturation at 100°C	0,0000	0,0000	0,0000	0,0015	0,0117	0,0000	0,0000		0,0000	0,0000
Cat's hair after heat denaturation at 100°C	0,0000	0,0000	0,0000	0,0015	0,0117	0,0000	0,0000	0,0000		0,0000
IgG after heat denaturation at 63°C	0,0000	0,0000	0,0000	0,0015	0,0117	0,0000	0,0000	0,0000	0,0000	

Tab. 1. Comparision of the fluorescence intensity quotients – red fraction's to yellow's – for the Congo red bound to polymers (protein, polisaccharide) and cells. The Wilcoxon signed-rank test. results not marked in bold demonstrate statistical differences p ≥0.05

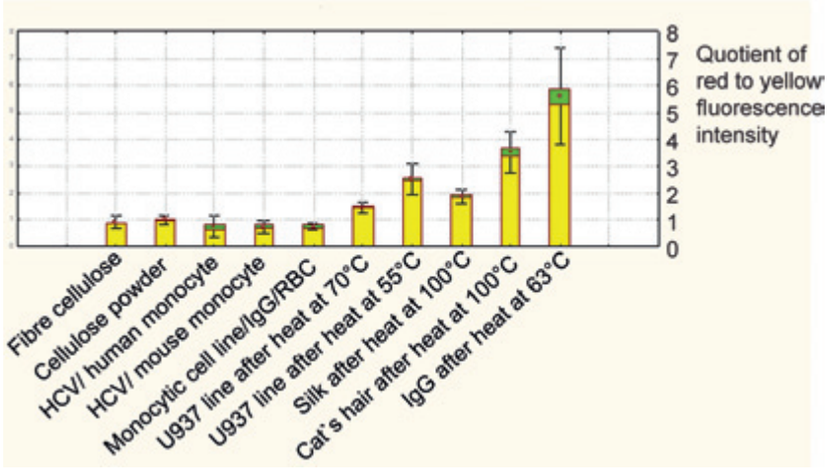


Fig. 2. Mean values and standard deviation for the quotient of red and yellow fluorescence intensity

statistical analysis (Statistica 7) of the results showed the identity of the colour in images coming from the same type of samples. This analysis of data revealed the significance of differences and similarities between samples (Table 1).

The Wilcoxon signed-rank test (Table 1) analysis indicates statistically significant similarity in the colour between cellulose (particularly in the form of fibres) and the natively cell structures, where monocyte cell lines (activated by cancer cells HCV29T or immunological complexes of heat-aggregated IgG or IgG antySRBC/SRBC) associated with Congo red dye. As shown in chart (Figure 2), the heat-aggregated IgG (with CR) demonstrates the highest intensity of the red fluorescence, while the native cellular systems (with monocyte/macrophage and CR) and samples with cellulose (with CR) show the lowest intensity. Heat-treated proteins in the presence of Congo red demonstrate a shift of fluorescence spectrum towards red; however, the shift is dependent on the type of protein used and the temperature of denaturation.

Congo red solutions

Fluorescence measurements of the Congo red solutions of increasing hydrophobicity indicate that the emission spectra, irrespective of the excitation wavelength (365 nm, 470 nm and 546 nm), demonstrate one maximum – e.g. at 600 nm (in propanol) and at 650 nm (in water), depending on the solvent used (Table 2).

Furthermore, fluorescence intensity distinctly increases alongside with the increase of solvent hydrophobicity (Figure 3). However, it must be noted that a very large increase of fluorescence for the excitation light at a wavelength 546 nm can be caused by a small difference between excitation and emission wavelengths.

Wavelength of Congo red fluorescence maximum λem (nm)			
Solvents	Excitation wavelenght λex = 365 nm	Excitation wavelenght λex = 470 nm	Excitation wavelenght λex = 546 nm
Water	631	623	636
Metnanol	603	601	605
Ethanol	603	600	603
1-propanol	594	597	604

Tab. 2. Wavelenght of Congo red fluorescence maximum λem (nm)

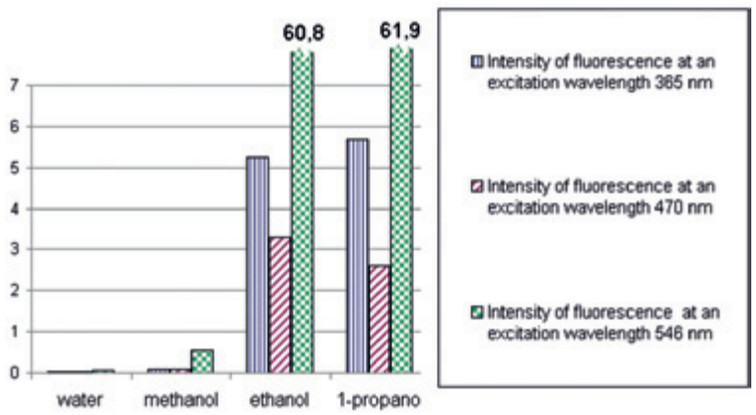


Fig. 3. Relativity between the intensity of Congo red's fluorescence and the solvent

Discussion

While comparing emissions of light by the Congo red solutions it was observed that the maximum of fluorescence shifts depending on the polarity of solvent (Table 2). Furthermore, the fluorescence intensity is positively correlated with hydrophobicity (Figure 3). A shift of the peak of emission towards longer wavelengths, as the hydrophobicity of the environment increases, is well documented (28). Unfortunately, despite an increase in fluorescence intensity for less polar solutions, it is still relatively low and occurs for a wavelength corresponding to yellow colour, which is in contrasts with observations for solid substances stained by CR (Figure 2). Movement of the emission peak towards red can be explained by investigations revealing that Congo red fluorescence shifts towards longer wavelengths and fluorescence intensity increases under the rising pressure (29). By comparing our observations with literature data we can conclude that CR binding to the polymer matrix may reveal a similar mechanism to the micelle's deformation mechanism under high pressure. Thus, the stronger the interaction between particles of dye and the ground, the bigger the shift of fluorescence towards longer wavelengths (red). This could explain the change of fluorescence colour depending on binding conditions (Table 1). In the case of association at a room temperature with cellulose and

surface structures of living cells binding of CR is limited to the surficial, properly shaped and easily available parts of polymers (proteins and polysaccharides). However, thermal modification (at 63°C) of immunoglobulin causes very large changes in the spatial structure of the protein. As a consequence, particles of CR penetrate protein structure very deeply forming a new super-molecule which consists of protein chains and molecules of CR. New molecule has a totally different spatial structure compared to the native protein. In the first case (surface interactions) the pressure of the ground that deforms Congo red micelle is relatively small, whereas in the other case, strong interaction with denatured protein chains causes large deformation of the micelle. For other samples, interactions between the micelle and the polymer matrix show intermediate values; this is indicated by the colour of emitted fluorescence (Figure 2). Spectrum of fluorescence can be explained by the fact that both silk protein and hair keratin do not substantially change their internal structure during heat-modification, as is the case of aggregated immunoglobulin. However, this connected with pressure phenomenon does not explain the reverse trend that occurs during heat denaturation of whole cells in the presence of CR. The transfer of fluorescence towards shorter wavelengths (yellow) in effect of the increasing temperature of modification can be caused by other factors present in a complex structure of the whole cell, which are not taken into consideration. During heat modification, depending on time and temperature of the process, CR is able to stain all structures, both surface (proteins, sugars and even lipid membrane) and internal (all cellular structures together with nucleic acid).

The analysis of the light emitted by Congo red associated with polymers such as simple models of pure proteins or polysaccharides, suggests that the spectrum of fluorescence is determined by the strength of interaction between polymers and particles of CR. More detailed analysis of all factors affect spectrum of fluorescence (especially in heterogeneous systems like whole cells), taking also into account the polarity of the environment, would allow to use this phenomenon as a convenient predictor of interaction strength between polymers and supramolecular dyes. Such assessment would influence the interpretation of results obtained during studies on the activity changes of protein structures (e.g. receptors) that were caused by supramolecular dyes, with particular reference to Congo red.

References:

1. Roterman I., No K.T., Piekarska B., Kaszuba J., Pawlicki R., Rybarska J., Konieczny L.: Bis azo dyes – studies on the mechanism of complex formation with IgG modulated by heating or antigen binding. *J. Physiol. Pharmacol.* 1993, 44(3): 213-232.
2. Stopa B., Konieczny L., Piekarska B., Roterman I., Rybarska J., Skowronek M.: Effect of self association of bis-ANS and bis-azo dyes on protein binding. *Biochimie.* 1997, 79(1): 23-26.
3. Stopa B., Piekarska B., Konieczny L., Rybarska J., Spolnik P., Zemanek G., Roterman I., Krol M.: The structure and protein binding of amyloid-specific dye reagents. *Acta Biochim. Pol.* 2003, 50(4): 1213-1227.
4. Glenner G.G., Eanes E.D., Page D.L.: The relation of the properties of Congo red-stained amyloid fibrils to the conformation. *J. Histochem. Cytochem.* 1972, 20(10): 821-826.
5. Roterman I., Krol M., Nowak M., Konieczny L., Rybarska J., Stopa B., Piekarska B., Zemanek G.: Why Congo red binding is specific for amyloid proteins – model studies and a computer analysis approach. *Med. Sci. Monit.* 2001, 7(4): 771-784.
6. Rybarska J., Konieczny L., Piekarska B., Stopa B., Roterman I.: The detection of specific acute phase serum protein complexes and immune complexes by Congo red binding. *J. Physiol. Pharmacol.* 1995, 46(2): 221-231.
7. Pepys M.B.: Pathogenesis, diagnosis and treatment of systemic amyloidosis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 2001, 356(1406): 203-210.
8. Kaszuba J., Konieczny L., Piekarska B., Roterman I., Rybarska J.: Bis-azo dyes interference with effector activation of antibodies. *J. Physiol. Pharmacol.* 1993, 44(3): 233-242.
9. Piekarska B., Roterman I., Rybarska J., Konieczny L., Kaszuba J.: The melting of native domain structure in effector activation of IgG studied by using Congo red as a specific probe. *J. Physiol. Pharmacol.* 1994, 45(1): 147-162.
10. Piekarska B., Rybarska J., Stopa B., Zemanek G., Krol M., Roterman I., Konieczny L.: Supramolecularity creates nonstandard protein ligands. *Acta Biochim. Pol.* 1999, 46(4): 841-851.
11. Rybarska J., Konieczny L., Roterman I., Piekarska B.: The effect of azo dyes on the formation of immune complexes. *Arch. Immunol. Ther. Exp. (Warsz.)* 1991, 39(3): 317-327.
12. Zemanek G., Rybarska J., Stopa B., Piekarska B., Spolnik P., Konieczny L., Roterman I., Bugajski A.: Protein distorsion-derived mechanism of signal discrimination in monocytes revealed using Congo red to stain activated cells. *Folia Histochem. Cytobiol.* 2003, 41(3): 113-124.
13. Ding F., Li N., Han B., Liu F., Zhang L., Sun Y.: The binding of C.I. acid red 2 to human serum albumin: Determination of binding mechanism and binding site using fluorescence spectroscopy. *Dyes and Pigments* 2009, 83(2): 249-257.
14. Yilmaz E., Memon S., Yilmaz M.: Removal of direct azo dyes and aromatic amines from aqueous solutions using two beta-cyclodextrin-based polymers. *J. Hazard. Mater.* 2010, 174(1-3): 592-597.
15. Lorenc-Grabowska E., Gryglewicz G.: Adsorption characteristics of Congo red on coal-based mesoporous activated carbon. *Dyes and Pigments* 2007, 74(1): 34-40.
16. Briggs C., Burgin S.: Congo red, an effective stain for revealing the chytrid fungus, batrachochytrium dendrobatidis, in epidermal skin scrapings from frogs. *Mycologist* 2004, 18(3): 98-103.
17. Nilsson K.P.R.: Small organic probes as amyloid specific ligands – past and recent molecular scaffolds. *FEBS Lett.* 2009, 583(16): 2593-2599.
18. Eisert R., Felau L., Brown L.R.: Methods for enhancing the accuracy and reproducibility of Congo red and thioflavin T assays. *Anal. Biochem.* 2006, 353(1): 144-146.
19. Klunk W.E., Debnath M.L., Pettegrew J.W.: Chrysamine-G binding to alzheimer and control brain: Autopsy study of

a new amyloid probe. *Neurobiol. Aging* 1995, 16(4): 541--548.
20. Klunk W.E., Pettegrew J.W., Abraham D.J.: Quantitative evaluation of Congo red binding to amyloid-like proteins with a beta-pleated sheet conformation. *J. Histochem. Cytochem.* 1989, 37(8): 1273-1281.
21. Cooper J.H.: Selective amyloid staining as a function of amyloid composition and structure. histochemical analysis of the alkaline Congo red, standardized toluidine blue, and iodine methods. *Lab. Invest.* 1974, 31(3): 232-238.
22. Glenner G.G., Eanes E.D., Page D.L.: The relation of the properties of Congo red-stained amyloid fibrils to the -conformation. *J. Histochem. Cytochem.* 1972, 20(10): 821-826.
23. Pigorsch E., Elhaddaoui A., Turrell S.: Spectroscopic study of pH and solvent effects on the structure of Congo red and its binding mechanism to amyloid-like proteins. *Spectrochim. Acta, Pt A: Mol. Spectrosc.* 1994, 50(12): 2145-2152.
24. Kowalczyk A.: Biospektroskopia. Twardowski J. (ed.); Vol. 3, PWN, Warszawa 1989, pp. 9-67.
25. Shukla A., Mukherjee S., Sharma S., Agrawal V., Radha Kishan K.V., Guptasarma P.: A novel UV laser-induced visible blue radiation from protein crystals and aggregates: Scattering artifacts or fluorescence transitions of peptide electrons delocalized through hydrogen bonding? *Arch. Biochem. Biophys.* 2004, 428(2): 144-153.
26. Linke R.P.: Highly sensitive diagnosis of amyloid and various amyloid syndromes using Congo red fluorescence. *Virchows Arch.* 2000, 436(5): 439-448.
27. Zhang Y., Xiang X., Mei P., Dai J., Zhang L., Liu Y.: Spectroscopic studies on the interaction of Congo red with bovine serum albumin. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 2009, 72(4): 907-914.
28. Nekipelova T.D., Shishkov V.S.: Solvent effect on the spectral characteristics and quantum yields of the photolysis of alkylated 1,2-dihydroquinolines. *High Energy Chemistry* 2004, 38(5): 315-322.
29. Wang Z.P., Zhang Z.M., Ding Z.J.: Photoluminescence study of Congo red molecules under high pressure. *J. Lumin.* 2007, 122(123): 237-240.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 25-30

EVALUATION OF AVAILABLE PROTEIN MASS SPECTRA PRE-PROCESSING ALGORITHMS

Michal Marczyk

Institute of Automatic Control, Silesian University of Technology, 44-100 Gliwice, Poland
Michal.Marczyk@polsl.pl

Abstract: Analysis of the profiles of body fluids by MALDI – ToF mass spectrometry is a popular technique used for searching of protein biomarkers that have potential application in early detection and diagnose of a cancer. A typical strategy used in validation of new biomarkers involves two basic aspects. Predictive properties of discovered differing features must be proven by classification of patients and controls spectra. Then, these features must be connected to proteins / peptides present in the analysed samples. Therefore, most mass spectra pre-processing procedures are based on reducing dimensionality of the spectrum to only significant features (peaks), assuming that each peak corresponds to a single protein / peptide and its position in the m/z scale, and height carry direct information on the composition of the test substance.
In the literature there are several different approaches of mass spectra pre-processing. But so far there are no standards for selection of techniques that are the most effective for this type of data. Only the pre-processing steps that should be done in order to extract the desired information from raw spectra are specified. This paper presents some algorithms that are compared on two levels. By the classification of real data sets differentiating potential of detected features was examined and the ability to reconstruct proteins / peptides in the test sample was checked. Since the composition of the specimen is not known, we used a virtual machine to generate artificial spectra.
Despite many published studies, scientists searching for biomarkers, still encounter many serious problems. Existing methods for mass spectra pre-processing are very sensitive to changes in data collection protocols, or instrumentation. Identified biomarkers of cancer vary between different research groups. The key is to choose the appropriate settings of the methods used. Thus, there is a need to test new procedures and automate the tuning of parameters of existing algorithms. Our simulations showed, that Align and CWT algorithms eliminates false positive peaks efficiently and that Align is the most flexible for changes in signal quality from all studied mass spectra pre-processing packages.
Keywords: spectra pre-processing, MALDI - TOF, biomarkers

Introduction

Due to the fact that proteins are involved in almost all important biological functions mass spectrometry (MS) as a technique that is commonly used in proteomics can be adapted to identify disease related patterns. Matrix-assisted laser desorption and ionization time of flight (MALDI-TOF) and its variant surface-enhanced laser desorption and ionization time of flight (SELDI-TOF) mass spectrometers can detect the components at a very low concentrations. This can help in the diagnosis, disease progression monitoring and response to treatment or prognosis of a cancer.

A mass spectrometer does not actually measure the molecular mass directly, but rather the mass-to-charge ratio (M/Z) of the ions formed from the molecules. The charge on an ion is denoted by the integer number *Z* of the fundamental unit of charge, and the mass-to-charge ratio M/Z therefore represents Dalton per fundamental unit of charge. In many cases, the ions encountered in mass spectrometry have just one charge (Z=1) so the M/Z value is numerically equal to the molecular (ionic) mass in Da. So, a mass spectrum is a graph of ion intensity as a function of mass to charge ratio. Hence, the mass spectrum of a sample is a pattern representing the distribution of components (atoms or molecules) by mass (more correctly: M/Z ratio) in a sample.

The first approaches to detect protein biomarkers using MS met with great optimism (Adam et al. 2002; Paweletz et al. 2001; Petricoin et al. 2002). Unfortunately, these results were soon challenged. It turns out, that the proteomic data obtained through the MS, contain a huge number of different types of noises and artefacts. Some researchers, however, saw this as a challenge to create a more sophisticated bioinformatics tools. During this time many papers that compare different pre-processing techniques appeared (Cruz-Marcelo et al. 2008; Emanuele II et al. 2009; Meuleman et al. 2008; Yang et al. 2009).

Nowadays, there are no standards for MS data pre-processing, but most of the algorithms realizes following steps: baseline subtraction, which removes systematic artefacts, usually attributed to clusters of ionized matrix molecules hitting the detector during early portions of the experiment or to detector overload; smoothing, which reduces the level of random noise, typically electronic or chemical; normalization, which corrects for systematic differences in the total amount of protein desorbed and ionized from the sample plate. The crucial step is a peak detection – the process of identifying locations on the M/Z scale that correspond to specific proteins or peptides striking the detector.

In this study we propose a comparison of four MS pre-processing algorithms that are commonly used and freely available. We checked differentiating potential of detected peaks using real data and the ability to reconstruct proteins/peptides in the test sample using virtual data.

Materials and methods

Data

Real data used in a study were obtained at the Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice Branch. 92 patients diagnosed with clinical stage I or II breast cancers were included in the study, of average age 58.5 years (range 31–74 years). Patients were classified according to the TNM scale; the majority were scored as T1 and T2 (47% and 45%, respectively) as well as N0 and N1 (75% and 24%, respectively), and none had diagnosed metastases (all M0). Serum samples were collected before the start of therapy. 104 female volunteers were included as a control group; they were required to be free of any known acute or chronic illness and were not treated with any anticancer therapy in the past. The average age in this group was 54 years (range 32–77 years). Samples were collected and processed following a standardized protocol and analyzed using an Autoflex MALDI-ToF mass spectrometer; the analyzer worked in the linear mode and positive ions were recorded in the mass range between 2,000–10,000 Da (Nowicka et al. 2008, Pietrowska et al. 2009).

Since the true composition of analysed samples is not known, it is difficult to create an unbiased comparison study. Following others (Yang et al. 2009, Cruz-Marcelo et al. 2008), we used a virtual mass spectrometer developed by Coombes et al. (2005), which is based on the physical principles underlying the instrument and allows us to generate realistic virtual spectra with a known underlying protein (peptide) composition. Composition of protein species in artificial spectra was chosen the same as that estimated from the real data.

Distribution of peaks across samples is quantified by such measures: its prevalence, defined by the proportion of samples in the population containing the protein; the mean and standard deviation of corresponding peak intensity across samples that contain the protein and the mean and standard deviation of the M/Z value across samples that contain the protein.

For each peak we first determined whether it is present in the sample by a random Bernoulli trial with probability defined by the value of the protein prevalence. We then generated peak's intensity by drawing a random number distributed log-normally (Morris et al. 2005) and peak's M/Z by drawing random number distributed normally. All the distributions parameters were fitted to real data to preserve the nature of protein spectra. Using these setting 50 datasets was created with 50 spectra each.

Algorithms

Align is a mass spectra pre-processing tool created in Matlab graphical user interface (Marczyk 2009). It estimates the baseline within multiple shifted windows with varying size and regress the baseline to the window points using a spline approximation. Noise is removed using variable span smoother based on local linear fits (Friedman 1984). Spectra are aligned using peak alignment by fast Fourier transform (Wong et al. 2005). To normalize spectra Total Ion Current method is used. Peak detection algorithm first finds peaks in the spectrum using the first derivative. Next it checks a ratio of maximum to left and right side minimum separately. Small amplitude peaks, where none of these ratios is greater than given threshold value, are deleted. Noise reduction is obtained by replacing similar intensity peaks, which are in close M/Z neighbourhood by one, highest peak. Peaks with intensity lower than a given threshold are also deleted.

LIMPIC is a computational method for the detection of protein peaks from linear-mode MALDI-TOF data (Mantini et al. 2007). A noise reduction technique is performed at first: the signal-to-noise ratio (SNR) is enhanced by using a smoothing procedure based on a Kaiser filter. Next, the resulting spectrum is used to sequentially estimate the baseline drift and the non-uniform noise level. Spectrum is divided into signal blocks and the blocks showing peaks are selected on the basis of the kurtosis. Peak detection procedure is as follows: if the point intensity is the highest among its nearest ± f points, a peak is detected in that position. The last step is the elimination of the detected peaks with a SNR lower than a given threshold.

Mass Spec Wavelet R package contain a continuous wavelet transform (CWT)-based peak detection algorithm that identifies peaks with different scales and amplitudes (Du et al. 2006). By transforming the spectrum into wavelet space, the pattern-matching problem is simplified and in addition provides a powerful technique for identifying and separating the signal from the spike noise and colored noise. This transformation, with the additional information provided by the 2DCWT coefficients greatly enhances the effective signal-to-noise ratio. With this technique baseline removal and smoothing are not required before peak detection.

Cromwell is a set of Matlab scripts implementing the methods for pre-processing of MS data developed by the bioinformatics group at the MD Anderson Cancer Center (Coombes et al. 2007; Morris et al. 2005). Smoothing is performed by wavelet regression using the undecimated discrete wavelet transform. The baseline is estimated by computing a monotone local minimum curve. Normalization is performed by scaling each individual spectrum so that the mean of its intensities is equal to 1. The algorithm for peak detection finds local maxima in the denoised, baseline subtracted spectrum and retains as peaks those with a signal-to-noise ratio greater than a user-defined threshold.

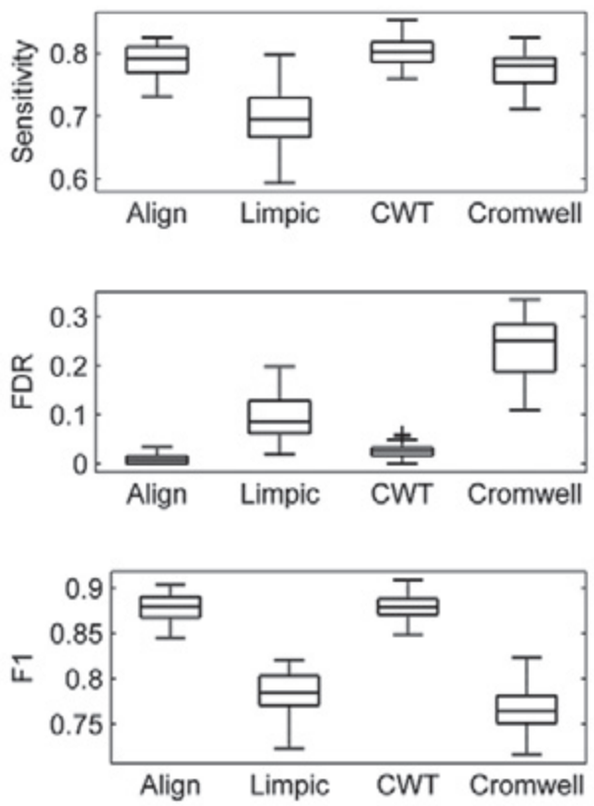


Fig. 1. Boxplots comparing peak detection results on 50 virtual datasets. Top – sensitivity, middle – FDR and bottom – F1.

Peak detection performance

In this comparison peak detection is performed by using mean spectrum. This procedure is commonly applied and has the following advantages (Morris et al. 2005): the noise in the mean spectrum is decreased; peak finding in the mean spectrum is more sensitive; small, but consistent peaks are easily seen in the mean spectrum; it allows finding peaks present in only few samples.

For each virtual dataset we calculated the following three measures: the sensitivity, which is the proportion of matched peaks, false discovery rate, which is the proportion of found peaks that were not matched and the F1 measure, introduced by Yang et al. (2009), as a compromise between false discovery rate and sensitivity. Detected peak was acknowledged as matched, if it was ± 0.3% M/Z away from known value of true peak.

In most cases the highest F1 is implied by the best (highest) sensitivity. That is why we tune the algorithms parameters to get the highest F1. Because we also wanted to check flexibility of the pre-processing algorithms in terms of signal change, optimal settings were fixed during all tests.

To determine the trade-off between sensitivity and FDR we estimated operating characteristics (OC) by varying significant parameters of each algorithm (Emanuele II et al. 2009). Strictly, for each parameter settings we calculated sensitivity and FDR for each dataset mean spectrum. By calculating mean value and error margin (assuming normal distribution of sensitivity and FDR) we obtained points on the OC plot with confidence intervals on both axes.

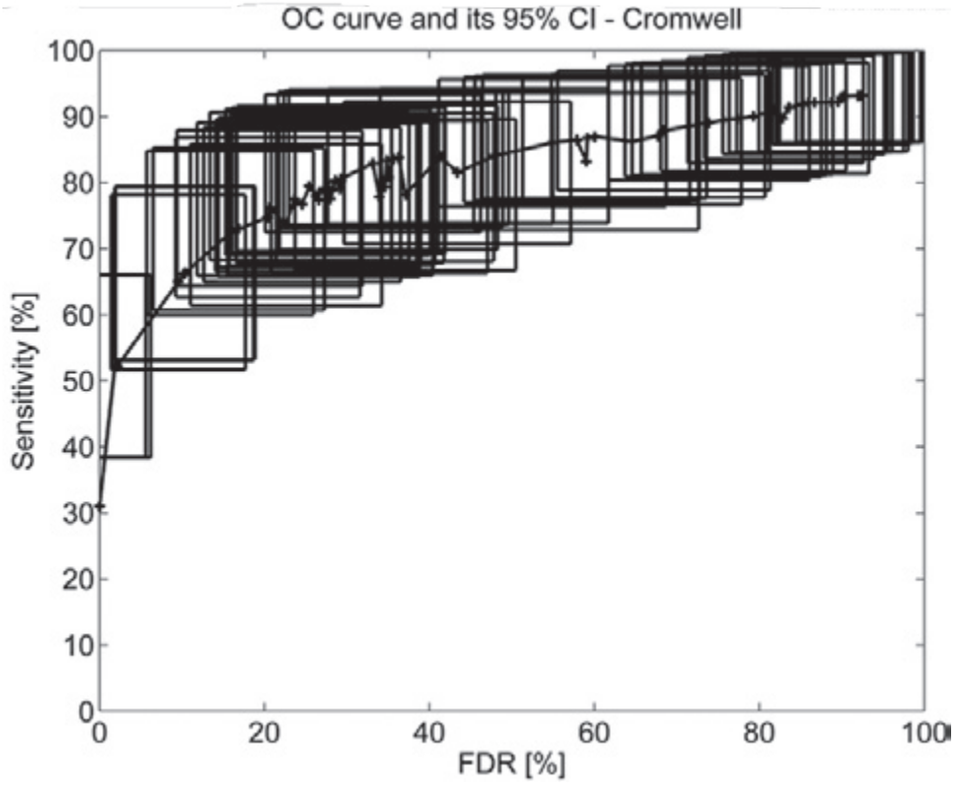


Fig. 2. Cromwell package operating characteristics and its 95% confidence intervals.

Classification performance

Differentiating potential of detected features was examined by the classification of real data using Support Vector Machine (SVM) (Vapnik 1995) and measured by classification error. SVM is a classifier that is widely used because of its generally good performance in complex classification problems and especially in applications, such as the problem being studied here, where the number of features is larger than the total number of samples. Transformation to a non-linear space is achieved by using a radial basis kernel function. To assess the performance of the classifier we used multiple random sampling validation with 500 iterations (Michiels et al. 2005). In each iteration test group was build using 50% of all samples. Because of the non normal distribution of peak's intensities, Mann Whitney U-test was used for feature selection. We presented here classification error for classifiers created from 1 to 40 features with lowest p-values. In each random sampling classification error is a maximum likelihood estimate of failure (if sample is classified incorrectly). By providing a large number of iterations, we can conclude that this estimate is approximately normally distributed.

Results

Extraction of representative peaks from noisy data measured by sensitivity, FDR and F1 is presented in fig. 1. Both Align and CWT procedures showed good efficiency, because despite high sensitivity, which brings ability to reconstruct content of testing sample, FDR is at really low level. It means that most of the detected peaks are significant, because they correspond to true proteins/peptides. Cromwell package gave quite good sensitivity, but because of big number of detected peaks, FDR value was at high, not acceptable level.

Fig. 2. presents OC curve with 95% confidence intervals for Cromwell. This tool gave the biggest range of FDR, however, confidence intervals, which width is similar in all algorithms, are too wide for clear results presentation. Fig 3. shows sensitivity and FDR in only chosen operating points, without confidence intervals and cut to FDR equal 30%, because in biomarker searching peaks detected at higher FDR values are useless. In the Align OC curve, at the beginning of the scale, some instability was observed. It is caused by too strict peak picking settings, which neglects most of the peaks. Despite that, Align and CWT algorithms can guarantee good sensitivity from quite low FDR values. To produce OC plots we varied parameters in a big range around optimal settings, so such results showed that these algorithms are quite flexible. Peaks detected after using Cromwell reached the highest sensitivity, but at high FDR level, which is not desirable behaviour. The smallest sensitivity in tested FDR range appeared in Limpic OC curve.

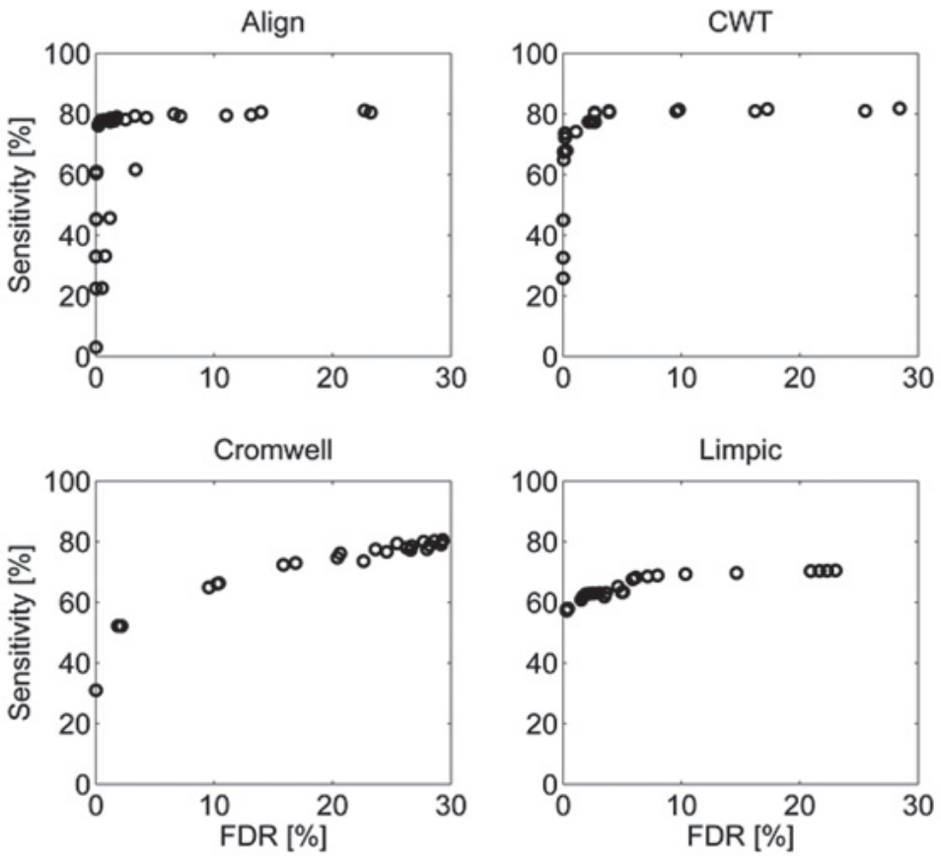


Fig. 3. Comparison of operating characteristics of all four programs.

In classification of protein spectra for biomarker discovery, the key issue is to find a small group of features with the highest differentiating abilities. So presented results (fig. 4) contain classification error for validation of classifiers built from 1 to only 40 features. In such comparison Align outperforms other algorithms, where minimal error was about 25% and the shape of the plot is preferable. Classification after using CWT and Cromwell gave minimal error around 35 % and Limpic – 38%. One of the reasons of such results is a fact that only in Align software, there is a spectra alignment pre-processing step introduced. Parameters tuning will for sure reduce the classification error, but it was stated, that they are fixed during all tests. These scores showed that Align is the most flexible for changes in signal shape and noise from all tested algorithms.

Conclusions

In this paper a comprehensive comparison of four mass spectra pre-processing algorithms was evaluated. Results showed that Align and CWT provides very good performance, it means good sensitivity with small FDR and nice OC, despite fact, that pre-processing methods implemented in these programs differ much. However, we can conclude some common remarks. In protein spectra most noises are non-linear, so using methods with fixed window size may fail. In Align both steps are conducted with varying window sizes, fitted to analysed data. CWT avoids the problem by multiscale smoothing and no need for baseline correction. Taking into account peak detection, it can be concluded that max/min ratio and intensity thresholding connected with noise peaks removal used in Align gave very robust and flexible combination. Concept of forming ridge lines in CWT also removes many false positive peaks. Last test proved that default parameters setting in Align software is good enough to remove most of the false positives, despite quality changes in analysed signals.

Funding: This work was partially supported by BW RAU1/2010.

References

1. Petricoin III E.F., Ardekani A.M., Hitt B.A., Levine P.J., Fusaro V.A., Steinberg S.M., Mills G.B., Simone C., Fishman D.A., Kohn E.C., Liotta L.A.: Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002, 359: 572--577.
2. Adam B.L., Qu Y., Davis J.W., Ward M.D., Clements M.A., Cazares L.H., Semmes O.J., Schellhammer P.F., Yasui Y., Feng Z., Wright G.L. Jr.: Serum protein fingerprinting coupled with a pattern-matching algorithm distinguished prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res.* 2002, 62: 3609-3614.

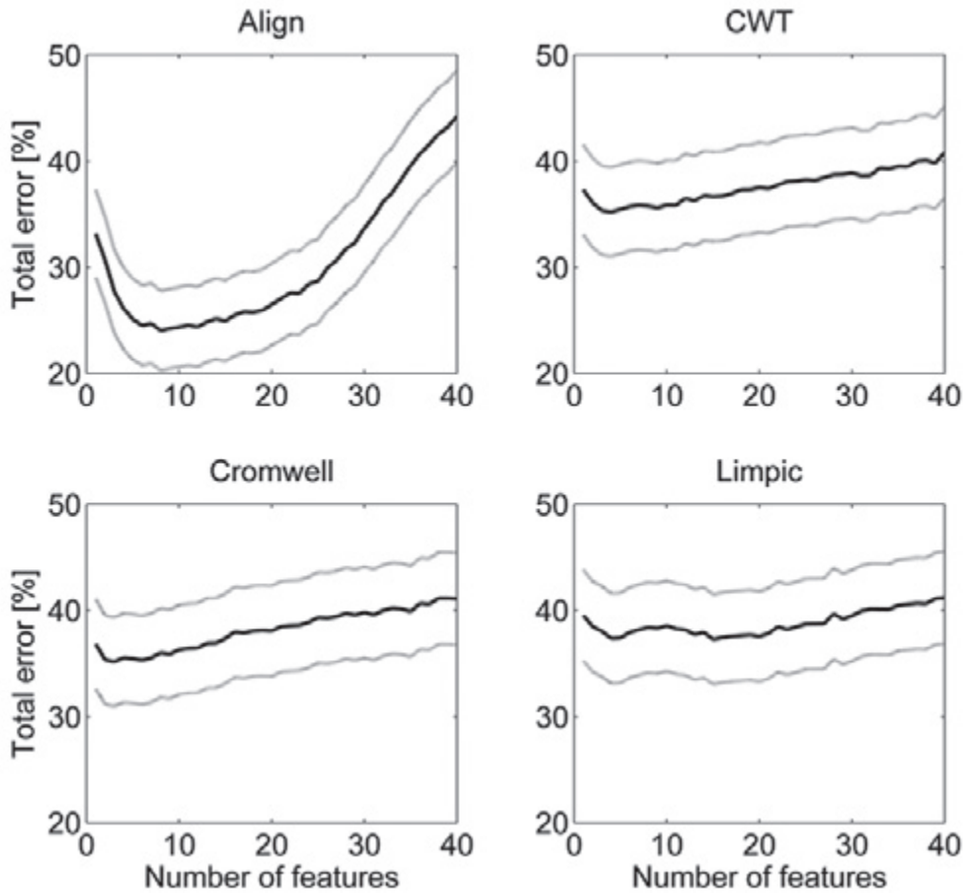


Fig. 4. Classification results of classifiers created from 1 to 40 most significant features mesured by total classification error.

3. Paweletz C.P., Trock B., Pennanen M., Tsangaris T., Magnant C., Liotta L.A., Petricoin III E.F.: Proteomic patterns of nipple aspirate fluids obtained by SELDI–TOF: potential for new biomarkers to aid in the diagnosis of breast cancer. *Dis. Markers* 2001, 17: 301-307.
4. Cruz-Marcelo A., Guerra R., Vannucci M., Li Y., Lau C.C., Man T.-K.: Comparison of algorithms for pre-processing of SELDI-TOF mass spectrometry data. *Bioinformatics* 2008, 24: 2129-2136.
5. Emanuele II V.A., Gurbaxani B.M.: Benchmarking currently available SELDI-TOF MS preprocessing techniques. *Proteomics* 2009, 9: 1754-1762.
6. Meuleman W., Engwegen J., Gast M., Beijnen J., Reinders M., Wessels L.: Comparison of normalisation methods for Surface-Enhanced Laser Desorption and Ionisation Time-Of-Flight Mass Spectrometry data. *BMC Bioinformatics* 2008, 9(88).
7. Yang C., He Z., Yu W.: Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. BMC Bioinformatics 2009, 10(4).
8. Coombes K.R., Koomen J., Baggerly K.A., Morris J.S., Kobayashi R.: Understanding the characteristics of mass spectrometry data through the use of simulation. *Cancer Informatics* 2005, 1: 41-52.
9. Nowicka E., Pietrowska M., Behrendt K., Walaszczyk A., Polanska J., Polanski A., Marczak L., Stobiecki M., Widlak P., Tarnawski R.: Potential clinical application of serum proteome mass spectrometry analyses in breast cancer patients diagnosis and management. *Eur. J. Cancer Suppl.* 2008, 6(7): 82.
10. Pietrowska M., Marczak L., Polańska J., Behrendt K., Nowicka E., Walaszczyk A., Chmura A., Deja R., Stobiecki M., Polański A., Tarnawski R., Widłak P.: Mass spectrometry-based serum proteome pattern analysis in molecular diagnostics of early stage breast cancer. *J. Transl. Med.* 2009, 7: 60.
11. Morris J.S., Coombes K.R., Koomen J., Baggerly K.A., Kobayashi R.: Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics* 2005, 21: 1764–1775.
12. Marczyk M.: Align – a software tool for mass spectra pre-processing. *Proceeding of the XI International PhD Workshop OWD* 2009, 88-91.
13. Friedman J.H.: A Variable Span Smoother, *LCS Tech. Rep.* 1984, 5.
14. Wong J.W.H., Durante C., Cartwright H.M.: Application of Fast Fourier Transform Cross-Correlation for the Alignment of Large Chromatographic and Spectral Datasets. *Anal. Chem.* 2005, 77: 5655-5661.
15. Mantini D., Petrucci F., Pieragostino D., Del Boccio P., Di Nicola M., Di Ilio C., Federici G., Sacchetta P., Comani S., Urbani A.: LIMPIC: a computational method for the separation of protein MALDI-TOF-MS signals from noise. *BMC Bioinformatics* 2007, 8: 101.
16. Du P., Kibbe W.A., Lin S.M.: Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* 2006, 22: 2059-2065.
17. Coombes K.R., Baggerly K.A., Morris J.S.: Pre-processing mass spectrometry data. In: Dubitzky W., Granzow M., Berrar D.P. (eds). New York: Springer 2007: 79-99.
18. Vapnik V.N.: The nature of statistical learning theory. New York: Springer-Verlag 1995.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 31-37

COMPARTMENTAL MODEL OF THE PHARMACOKINETICS OF DRUGS EXCRETED BY BILE OR BY NON-RENAL MODE

Jan Pawel Jastrzebski[1], Rafal Urniaz[1]

Department of Plant Physiology and Biotechnology Faculty of Biology, The University of Warmia and Mazury in Olsztyn, ul. Oczapowskiego 1A/113, 10-719 Olsztyn/Poland, bioinformatyka@gmail.com (Jan Pawel Jastrzebski); sitcome21@gmail.com (Rafal Urniaz); tel: 00 48 89 523 4292; fax: 00 48 89 523 4881

Abstract: The theoretical biology analysis, uses mathematical and physical models and bioinformatics tools in practical medicine. The analysis *in silico* reduces time and costs of new designed drugs and provides wider spectrum of information for exploration. Moreover, bioinformatics analysis of genome and proteome, especially with the use of microarray allows specifying diagnosis and therapy. The methods enabling precise preparation of therapy by means of accurate medicine dosage are being currently searched, such an approach permits to increase the effectiveness of therapy, eradicate side effects (i.e. medicine overdosing), reduce the total cost of therapy and cut down its interval. The elaborated model involves the partition of human body into compartments having different physical features: weight and volume, or physiological character: metabolic efficiency, penetration rate of drug. Additional parameters can be introduced to the model: $W_{a/b}$, W_{wzb} and λ. Due to the nature of simulation each individual case can be run with high precision. The model is implemented in the GROW_4 virtual environment which is accessible free of charge on the websites: www.sitcome.vacau.com or www.uwm.edu.pl/bioinfo (in science/projects/grow).
Keywords: bioinformatics, modeling, algorithms, pharmacokinetic, codeine, bile mode, non-renal mode, compartment

Introduction

More often bioinformatic tools are used in medical analysis in accurate examination of an individual patient. Medical databases, computer analysis of microarray data or particular genetic analysis are developing in an intensive way in connection with a need to individualize patients. It is known that identical medicine dosage or dosage depending on weight only do not give identical effects for particular patients. Insufficient medicine dosage for one patient can be toxic in another case. It depends on different physical parameters of an organism such as weight or blood volume and individual physiological features. Medicine in an organism can be temporarily accumulated in organs causing pathological changes and systemic poisoning. Medicine accumulation can be prevented predicting if and when medicine concentration exceeds therapeutic dose and starts to be toxic.

The model of medicine distribution in organism presented in the following paper provides that each tissue is treated as a singular compartment, where the fate of the drug can run in a different way independently. Each tissue has attributed physical features (i.e. volume), physiological features (i.e. permeability of the drug, the drug half-life in tissue), which influence an overall distribution of a medicine in an organism. Functions and parameters of particular tissues and organs are changeable which allows individualization of simulation and adjusting the cycle of therapy to a single patient. Introducing additional parameters, such as distribution coefficient, coefficient of mass and drug binding allow more accurate individualization of therapy. Still, they require previous clinical testing of each patient.

The model is designed for distribution of codeine on the basis of medical data obtained from the analysis of the changes of medicine concentration in blood for 180 minutes. This data refers to 2 patients (A and B) [7]. The models for other tissues have been designed analogically for blood and verified comparing with statistical significance of theoretical models with empirical data. The preliminary of these data were presented elsewhere (section 3).

Material and methods

Increasing number of therapeutic substances is growing more and more frequently discussed issue about selection of an appropriate dose dependent on the individual characteristics of the person [1, 2, 3].

The presented model involves existence of 5 independent spaces in an organism which are separated by tissue barriers, limiting free distribution of therapeutic substance. They are: small intestine, liver, bloodstream, brain and bile (Fig. 1). Between those spaces therapeutic substance: flows (in all of spaces), is temporary accumulated (bile, bloodstream) or degraded and excreted (effects of liver's metabolism and brain degradation)

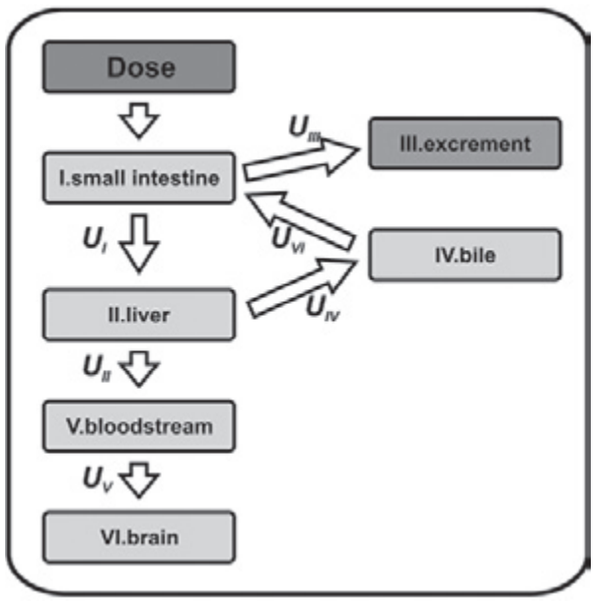


Fig. 1. Visualization of the compartmental model describing transport of the codeine in modeled organs. Arrows show the direction of codeine flow to next spaces (rectangles colored grey). U_I, U_{II}, U_{III}, U_{IV}, U_V, U_{VI} – speed of codeine transfer between spaces.

The model implies that the factors conditioning the dynamics and efficiency of distribution are among others: rate of blood flow through the organs, degree of perfusion and metabolism (degradation) of the therapeutic substance during its flow between spaces. Therapeutic substance transfer between separated spaces depends on speed of transfer (equation 1). It informs what quantity of substance is transferred to another space during the time. In elimination space (brain), the speed of transfer is convert to speed of elimination (E) and defined by the same way.

$$\upsilon_n = \frac{\Delta M}{\Delta t} \tag{1}$$

Many times, physiological conditions depends of structure, specification and activity of the organ. Substance is divided with no proportional quantum to the another organ and cells surrounding it [4]. Solution of this problem is addict to the equation a coefficient of separation „$W_{a/b}$" (equation 2), which means what part of primary dose (M_n) will be transfer to the current space (a or b).

$$\forall_{N_a, N_b \in R}\ N_a + N_b = 1 \qquad W_{a/b} = \frac{N_a}{N_b} \tag{2}$$

Liver's conversion a codeine to morphine is catalyzed by cytochrome P450 of enzyme CYP2D6. About 6-10% of European, 2% of Asian and 1% of Arabic populations [8] have weak function of CYP2D6. Codeine used to therapy for these patients doesn't have correct efficiency and they could feel pain even theoretically enough in-sensibilization [9]. Presented model contains two high-specific parameters: coefficient of liver's metabolism efficiency W_{wmw} and coefficient of intestine absorption W_{waj}, which give better correlation with experimental data. Main parameter describes medicament distribution in organism is actual mass of substance in the organ. The presented model used mathematical relations to describe more simply process of distribution, in similar way as it take place in organism. Equations describe quantity (mass) of substance in hypothetical organism space after time Δt are root of presented model. The general form of equation (3) describes quantitative relation between mass of substance (M). It goes to the next spaces with rate of transfer (υ), in time (Δt).

$$\Delta M_n = M_{avb} - \upsilon_n \cdot (\Delta t - \tau_{n-1}) \tag{3}$$

Normally physiological processes in organism (for example: drugs metabolism) are initiated with some delay compare to dose time (t_0), is necessary to introduce delay parameter (τ), which is a needful time to initialize the process [5]. The half-life time ($t_{1/2}$) (equation 4) means after what period of time, half dose of medicament will be eliminated from patient's organism. This is a key parameter to compare effects in organism for different groups of substances and it indicates how to basic dose should be select.

$$t_{1/2} = \frac{\ln\left(\frac{M_{max}}{\upsilon_n} \cdot 100\right)}{2} \tag{4}$$

Parameter M_{cN} means total mass of medicament available for modeled space. It is possible to calculate time used to transfer substance trough the biological membranes and in conclusion to calculate total time of substance resistant (t_{cN}) (equation 5). Total time of substance resistant and total mass of medicament in modeled organ inform us about therapeutic or toxic character of substance accumulation.

$$t_{cN} = \frac{M_{cN}}{\vartheta_n} \tag{5}$$

Mathematical series (equation 6) formed from semi-single equations (equation 5) and delays (equation 3), has a local extreme in t_C point. t_C point to the value is equal to total time of substance resistant in organism.

$$t_C = \sum_{n=1}^{n} \frac{M_{cN_n}}{\vartheta_n} + \tau_{n-1} \tag{6}$$

Changing of concentration (ΔC_n) in a better way presents the medicament's behavior in organism, because is possible to consider the volume of organ (V_n). This modification is important in pharmacokinetic point of view, because not mass, but medicaments concentration determines the therapeutic or toxic

substance character for tissue [6]. Equation 7 makes possible conversion of actual mass of medicament to concentration.

$$\Delta C_n = \frac{\upsilon_n \cdot (\Delta t - \tau_{n-1})}{V_n} \quad (7)$$

Because volumes of the same organs in different individuals are many times different (for example volume of blood in 6 years old girl and her mother) to correlate this differences with individual mass was added factor of multitude λ (equation 8) which means quotient of patient mass (M_{os}) and volume of the organ (eg. blood).

$$\lambda = \frac{M_{cs}}{V_n} \Rightarrow V_n \frac{M_{cs}}{\lambda} \quad (8)$$

To all of semi-equations was added a "p" parameter defined by condition nine, which prevents the negative values. This is an important part because changes (increases and decreases) the quantity of mass can not take negative values, it is physically impossible.

$$\forall\, \tau_a, \in N \qquad \Delta t - \sum_{n=1}^{n} \tau_n < 0 \Rightarrow p = 0 \quad (9)$$

For example, the equation 10 describes the changes of drug concentration in the blood. (Numbering is consistent with Figure 1)

$$p \cdot \Delta M_I - (M_D - \cdot \upsilon_I \cdot \Delta t) + \Delta M_{IV \to I} \quad (10)$$

$$p \cdot \Delta M_{II} = (W_{I/II} - (M_D - \Delta M_I)) - \upsilon_I \cdot (\Delta t - \tau_I) \quad (11)$$

$$p \cdot \Delta M_V = \Delta M_I - \upsilon_V \cdot (\Delta t - \tau_I - \tau_{III}) \quad (12)$$

$$\Delta C_V = \frac{\Delta M_V}{V_V} \quad (13)$$

The presented model was implemented in a specially built application called GROW_4. Experimental data, taken from the database at the Norwegian Institute of Public Health [7], experimentally determined codeine distribution in blood [16, 17] and statistical tests, such as a normality test (Shapiro-Wilk) and analysis of variance (one-way ANOVA) have been used to evaluate the model. In addition, the calculated data from the Grow_4 and PharmaCalc[18] programs has been compared.

The simulation results for a few standard therapeutic doses *per os* (M_d = 20mg (fig. 2); M_d = 15mg (dose A, fig. 3) and M_d = 30mg (dose B, fig. 4)[8,16,17] (tab. 3 and tab. 4)) have been compared in this paper. Additionally, the simulation of a 20mg dose was done for two types of patients:

– Patient A – European male with low efficiency of codeine conversion by CYP2D6 and
– Patient B – Asian male with high efficiency of codeine conversion by CYP2D6 [9].

The average laboratory results for the group of adult man who are healthy and don't smoke have been used to validate the designed model. The mean age of subjects ware 30.5 (range 25-36 years) and their weight 81.2 +/-8.6kg. The simulation parameters were obtained based on this information. The model contains the two types of parameters: high-individual (tab. 1) and low-individual (tab. 2) parameters. Individual parameters were developed to minimize differences between simulation results and experimental data using relative error (tab. 3, 4) and also to make the model more specific for each individual person. The low-individual parameters describe only the main pharmacokinetic trend of distribution and they are universal in predicting the concentration amoung the general human population. Conversely, to make person dependent simulations, the high-individual parameters were developed. High-individual parameters describe person dependent parameters like cofactors of absorption, cofactors of binding with blood proteins or cofactors of liver metabolism efficiency.

Parameters	Dose A	Dose B
M_d [mg]	15	30
Mass of organism [Kg]	81,2	
I*	10.9	
$W_{2/3}$*	75	
$W_{5/6}$*	95	
W_{waj}* [%]	20	18,33
W_{wzb}** [%]	7	
W_{wmw}* [%]	88	

Tab. 1. The high-individual parameters used in a computer simulations. * – parameters developed or estimated for the model. ** – parameters taken from literature [11], Md – basic dose, W2/3 – coefficient of separation, fraction of basic dose transferred to 2nd and 3th model space, W5/6 – coefficient of separation, fraction of basic dose transferred to 5th and 6th model space, W_{waj} – coefficient of drug absorbtion in intestine, W_{wzb} – coefficient of protein binding in blood, W_{wmw} – coefficient of metabolic efficiency in liver.

Parameters	Dose A	Dose B
u_I [mg/h]	7,25	
u_{II} [mg/h]	5,12	
u_{III} [mg/h]	4	
u_{IV} [mg/h]	0,1	
u_V [mg/h]	3,533	6
E [mg/h]	0,5	

Tab. 2. The low-individual model parameters used in computer simulations. Dose A – 15 mg ,Dose B – 30mg. U_I, U_{II}, U_{III}, U_{IV}, U_V, U_{VI} – speed of codeine transfer between spaces. t_I, t_{II}, t_{III}, t_{IV}, t_V, t_{VI} – delay required for transfer from one space to another.

Parameters	Dose A	Dose B
t_I [h]	0,13	
t_{II} [h]	0,3	
t_{III} [h]	1	
t_{IV} [h]	0,37	
t_V [h]	0,17	
t_{VI} [h]	0,79	

Presented data series (150 simulations) was exposed to statistical analysis for compliance with the normal distribution (Shapiro-Wilk test) and analysis of variance (one-way ANOVA) using Statistica 9th (Tab. 5).

Results

As shown in Fig. 2 model has a good correlation with experimental data and explains the differences between medicament's metabolism in outlying human population (European and Asian males). As we can see the small difference in metabolism efficiency can have a radical result in medicaments temporal commutation in organism.

The graph shows a good correlation between experimental data [11] (circles) and predicted data (Patient A (diamonds) – European male; Patient B (triangles) – Asian male); in regards to the codeine concentration in blood. This graph presents the data for a dose of 20mg of drug

Simulation results of the Grow_4 model for 15mg and 30mg dose (shown in Fig. 3 and Fig. 4) have a good correlation with experimental data in both cases. PharmaCalc model has a good correlation with only one of these doses – 15mg. The half-time elimination (Tab.3) in both cases are predicted well and values are in a middle range of literature results. Half-times calculated using the model (Tab. 3) for doses A and B have the same set of values as results taken from literature [11].

Assuming the average experimental results as a reference form model, we are able to compare and demonstrate prediction errors. As we can see on table 4, also in this time, the presented model has better result in comparison to PharmaCalc.

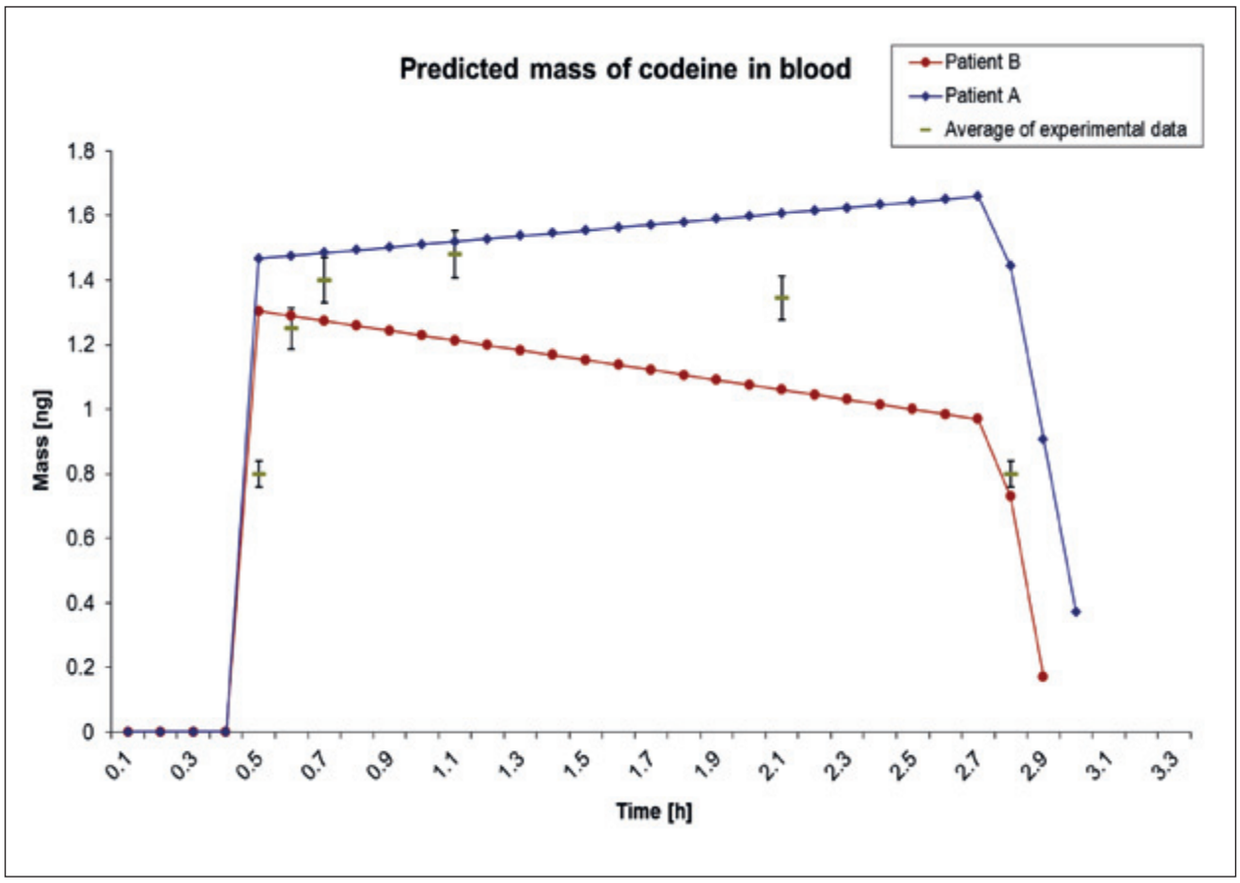


Fig. 2. Model correlation with experimental data of codeine elimination in the blood.

Moreover, the computer simulations can give much more results that values taken from literature or from experiments with the same statistical precision (Fig. 3, 4). This fact and its statistical significance (Tab. 5) give a high probability that the model can be a good tool to predict positive and individual dose for each patient to whom the high specific parameters were set.

However the accuracy of estimations using the presented model is significantly higher than for the PharmaCalc model (Tab. 4).

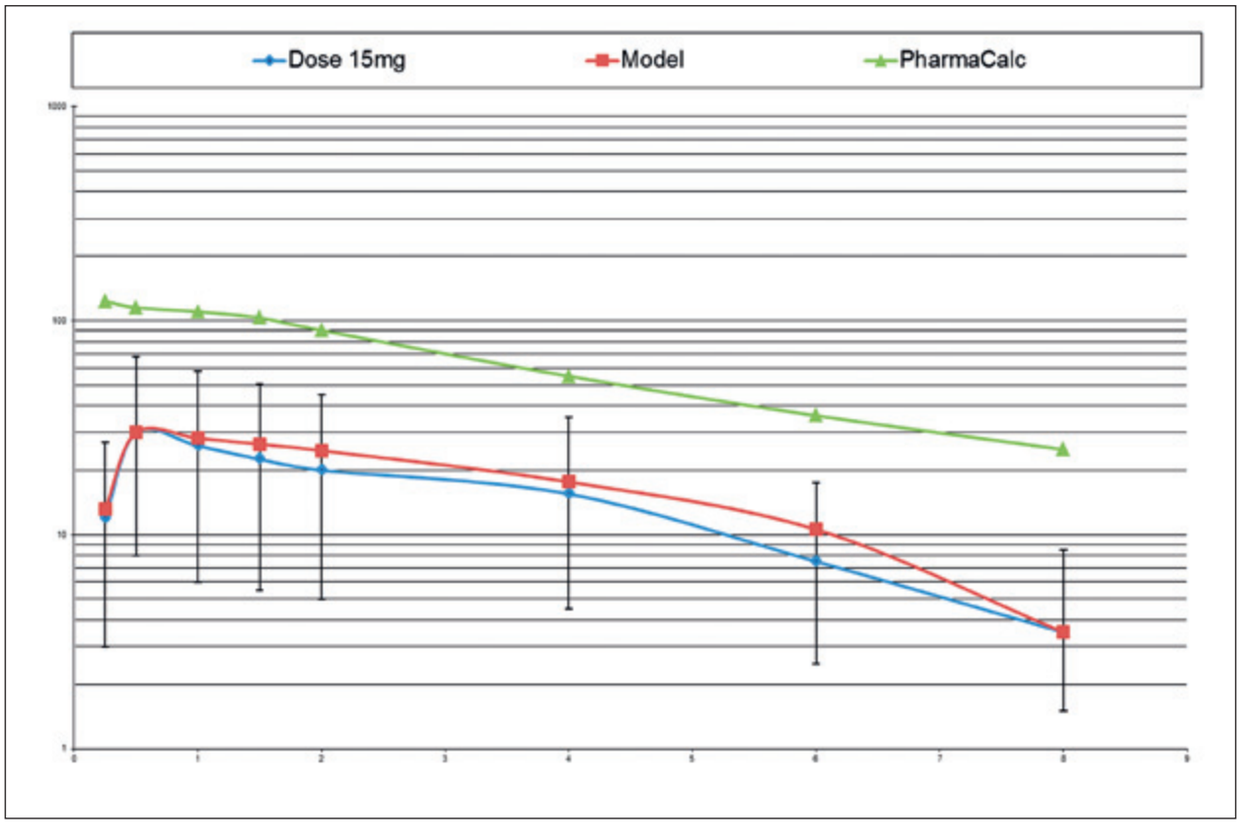


Fig. 3. Model correlation with experimental data of codeine elimination in the blood. Full field diamonds – experimental data for dose 15 mg [16, 17]; full field rectangles – predicted Grow_4 model data; full field triangle – predicted PharmaCalc data.

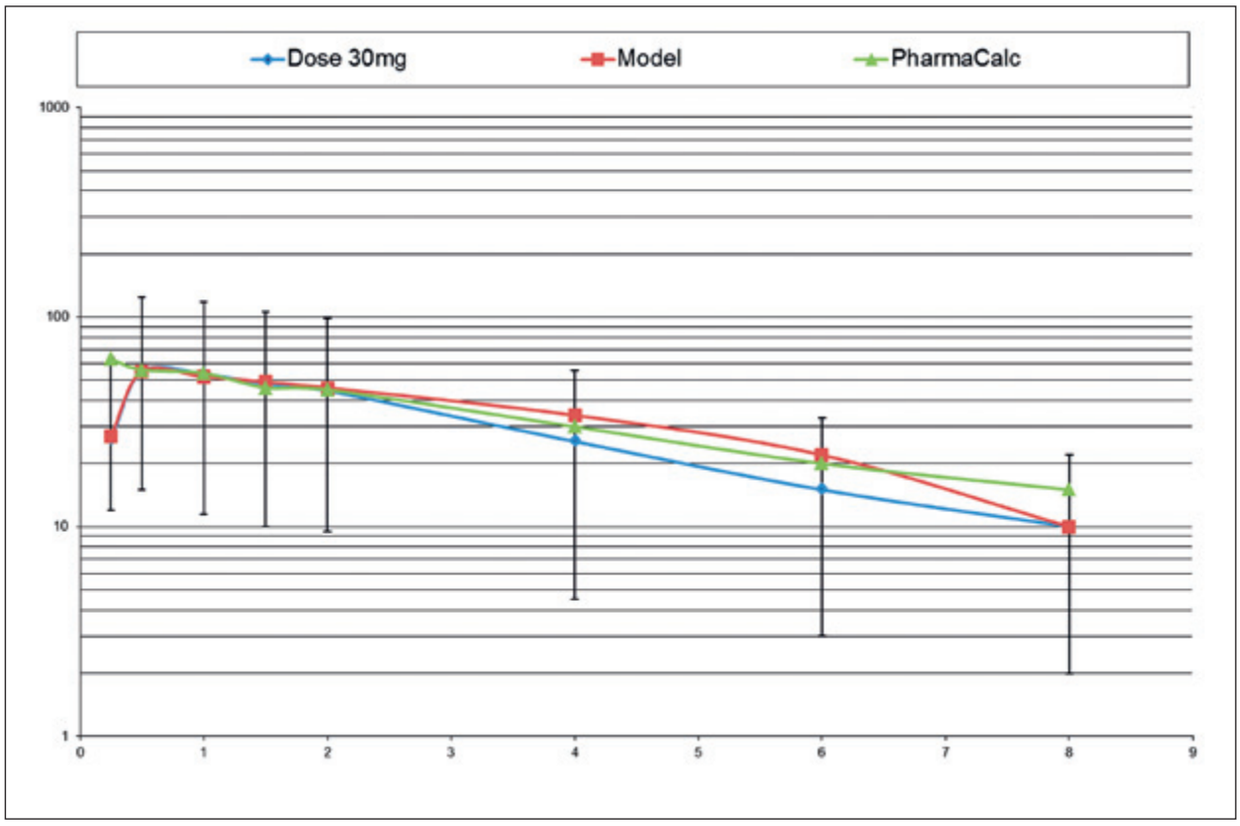


Fig. 4. Model correlation with experimental data of codeine elimination in the blood. Full field diamonds – experimental data for dose 30mg [16, 17]; full field rectangles – predicted Grow_4 model data; full field triangle – predicted PharmaCalc data.

	Literature [h]	Model [h]	PharmaCalc
Dose A	2,3 – 5,6	3,37	3,30
Dose B		3,41	3,30

Tab. 3. Comparison of half-life time ($T_{1/2}$) taken from literature [11] and predicted by models Grow_4 and PharmaCalc for two different dose quantity. Dose A – 15mg; Dose B – 30mg.

	Model [%]	PharmaCalc [%]
Dose A	12.63	379.56
Dose B	6.28	18.56

Tab. 4. Error measurement of predicted concentration of codeine in the blood for Grow_4 and PharmaCalc models in stand of experimental data. Dose A – 15mg; Dose B – 30mg.

Seria number	I	II	III	IV	V
Shapiro-Wilk test	0.963	0.969	0.960	0.948	0.975
One-way Anova	0.930	0.925	0.934	0.911	0.899
Results statistically significant?	yes	yes	yes	yes	yes

Tab. 5. Averaged results of the tests for hundred fifty simulations. Both used tests proofs statistical significance of obtained theoretical data. Numbers are assigned to spaces.

Discussion

The model presented in this paper has been compared with the two different virtual pharmacokinetics environments: R Package with PKfit [19] module and PharmaCalc [18]. The PKfit module has been designed to fit en experimental data to previously prepared compartmental models. The present model is a simulator which estimates the drug concentration and in blood of an individual patient. The changes of drug concentration are estimated based on high-individual parameters. PharmaCalc is the most similar already published simulator to the presented model. Therefore, this simulator has been chosen to prepare the accuracy tests.

The presented model has been tested on published medical data [16, 17] and compared to published pharmacokinetics models [18]. The data concerns the changes of drug concentration in blood [7, 11, 12, 13, 14, 15]. This kind of analysis for other tissues has not been published yet [9]. The significant level of compatibility between empirical and theoretical data has been statistically defined and presented in this work. However, the real significance of *in silico* data should be verified and confirmed by laboratory analysis and more clinical data. The application of this simulation tool for the therapy of patients, whose calculated dosage is almost identical to the standard dosage, can be the first step in clinical tests with the designed algorithm.

$W_{a/b}$, W_{wzb} and λ factors have been designed especially for this model. It allows a precise determination of simulation parameters for an individual case, for example an individual patient. These parameters truly and essentially influence a drug distribution in an organism, but are not measured in clinical practice. Therefore, elaborating the methodology to an empirical estimation of these factors should be a key stage to the clinical application of this model.

With the increase of clinical data we are planning to prepare a data base of individual models of therapy in cooperation with the GROW simulator. Based on the clinical data collected in this database it will be possible to prepare more, even theoretical drug distribution models for different types of organisms (different age, sex, diseases, residence or even basing on the genetic tests). This database allows to set simulator input data without most preliminary medical tests.

Acknowledgments

Authors are grateful for Professor Jan Sedzik for technical support.

References

1. Luo G., Johnson S., Hsueh M.M., et al.: In silico prediction of biliary excretion of drugs in rats based on physicochemical properties. *Drug Metab. Dispos.* 2010, 38: 422-430.
2. Poirier A., Funk C., Scherrmann J.M., et al.: Mechanistic modeling of hepatic transport from cells to whole body: application to napsagatran and fexofenadine. *Mol. Pharm.* 2009, 6: 1716-1733.
3. Fagerholm U.: Prediction of human pharmacokinetics-biliary and intestinal clearance and enterohepatic circulation. *J. Pharm. Pharmacol.* 2008, 60: 535-542.
4. Borodulin-Nadzieja L.: Human physiology – a handbook for students, Medical Publisher Gornicki. Wroclaw 2005.
5. Hermann T.: Pharmacokinetics: Theory and Practice, Medical Publisher PZWL: Warsaw 2002.
6. Kostowski W., Herman Z.: Pharmacology – Basics of pharmacology, Medical Publisher PZWL: Warsaw 2008.
7. Bachs L., Bramness J., Skurtveit S., et al.: Morphine Blood Concentration and Clinical Impairment in the Population of Drugged Drivers, www.icadts.org/T2004/pdfs/O9.pdf (07.2010).
8. Rossi S.: Het Australische Handboek van Geneesmiddelen, ISBN 0-9757919-6-7, 2008.
9. Srinivasan V., Wielbo D., Tebbett I. R.: Analgesic effects of codeine-6-glucuronide after intravenous administration. *European Journal of Pain* 1997, Volume 1, 3: 185-190.
10. Ceder G., Jones A.: Concentration Ratios of Morphine to Codeine in Blood of Impaired Drivers as Evidence of Heroin Use and not Medication with Codeine. *Clinical Chemistry* 1980-1984, 2001, 47: 11.
11. Panas M.: Pharmacokinetic and toxicokinetic parameters of some relevant drugs of abuse. *Przegląd Lekarski* 2001, (58)4: 373-375.

12. Bianchi G., Ferretti P., Recchia M., et al.: Morphine tissue levels and reduction of gastrointestinal transit in rats. Correlation supports primary action site in the gut. *Gastroenterology* 1983, 85(4): 852-858.
13. Ceder G., Jones A.W.: Concentration Ratios of Morphine to Codeine in Blood of Impaired Drivers as Evidence of Heroin Use and not Medication with Codeine. *Clinical Chemistry* 2001, 47: 1980-1984.
14. Gjerde H., Mordal J., Christophersen A.S., et al.: Comparison of drug concentrations in blood and oral fluid collected with the Intercept sampling device. *J. Anal. Toxicol*. 2010, 34(4): 204-209.
15. Boström E., Hammarlund-Udenaes M., Simonsson U.S.: Blood-brain barrier transport helps to explain discrepancies in in vivo potency between oxycodone and morphine. *Anesthesiology* 2008, 108(3): 495-505.
16. Quiding H., Anderson P., Bondesson U., et al.: Plasma concentrations of codeine and its metabolite, morphine, after single and repeated oral administration. *Eur. J. Clin. Pharmacol.* 1986, 30(6): 673-677.
17. Findlay J.W., Fowle A.S., Butz R.F., et al.: Comparative disposition of codeine and pholcodine in man after single oral doses. *Br. J. Clin. Pharmacol.* 1986, 22(1): 61-71.
18. http://www.pharma.ethz.ch/institute_groups/biopharmacy/teaching/pharma_calc.
19. Lee C., Lee Y., Setzer R.W.: A pharmacokinetic program (PKfit) for R, Annual Meeting of the Pharmaceutical Society of the Republic of China, Chia Nan University of Pharmacy and Science: Taiwan. 2005.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 39-45

THE BRAUN MACHINE AS A CLASSIFIER BASED ON ELEMENTARY PHYSICAL PHENOMENA IN THE CLASSIFICATION OF MULTIVARIATE DATA SETS

Mariusz Święcicki

Cracow University of Technology, Institute of Computer Science, Cracow, Poland
mswiecic@pk.edu.pl

Abstract: This article was presented to model the classifier algorithm based on elementary physical phenomena. The model for this is the result of research conducted by the author on the use of the mechanisms of physical phenomena in the classification of multidimensional data. In this article we have presented a model used for a classification of multidimensional data in a broader sense, called Braun's cathode machine. The internal structure of the machine presented on this paper has been based on the architecture of a cathode-ray tube – Braun's tube. For a machine model described this way a machine training algorithm has been proposed as well as response computing algorithms. In the final chapter we have presented the results of the machine tests for the notions connected with the classification and self-organization of multidimensional data.
Keywords: Braun's cathode machine, Data mining, Computational intelligence

Introduction

In recent years, there has been an increase of interest in biology as a source of inspiration to tackle many computational problems in the scope of data processing, artificial intelligence and optimization. This interest has been and is motivated by the willingness to select mechanisms that are used by natural systems and the attempt to adjust them to efficient problem solving in the aforementioned areas. An example of an effective application of natural mechanisms could currently be artificial neural networks, genetic algorithms, evolutionary algorithms which are being widely used in many fields of interest nowadays. The attempts to use elementary physical phenomena in data processing can be an alternative way of searching for natural computational models in relation to biology.

This article presents a data processing mechanism that uses the phenomena occurring during movements of particles in the electric field. The aforementioned occurrences have already been widely used in construction of Braun's cathode ray tube. In the first part of the article physical rights will be presented which have been used in data processing by a processing element called Braun's cathode ray machine or Braun's machine. In the second part of the article we will discuss the elements of the aforementioned processing element and its functioning mechanism will be presented. In the second part of this article we will also present Braun's cathode machine training algorithm and an algorithm which computes a response. The third part of this article presents the results of the computations carried out with the use of this model.

Physical phenomena used by Braun's machine

In this part of the article we will discuss physical phenomena which will constitute a functioning base for Braun's machine. The reaction of electric fields to charged particles is used in two aspects. First of all, it is used to accelerate or curb particle movements, and secondly, to change the direction of moving particles.

Speeding up the particles

May a moving particle have a m mass, *q* charge and *v0* velocity. Let us consider its motion in homogenous electrostatic field with *E* strength. If the direction of *v0* velocity is in accordance with the direction of electric field lines, this field operates on a charged particle with force *F=qE* which is constant so the movement of the particle is a uniformly variable motion. The charge will then accelerate or curb depending on the senses of vectors *v0* and *E*.

May a field in *A* point have a potential V_A and in *B* point – a potential V_B, so electrostatic potential energy of charge *q* in point *A* equals qV_A, and in *B* point, it equals qV_B. Appropriate kinetic energies are: in point *A* it is $(mv_0^2)/2$, in point B $(mv^2)/2$ where the *v* variable is a searched value. Taking advantage of the energy conservation law we obtain:

$$qV_A + \frac{mv_0^2}{2} = qV_b + \frac{mv^2}{2} \tag{1}$$

$$v = \sqrt{v_0^2 + \frac{2q}{m}U} = \sqrt{v_0^2 + \frac{2qEd}{m}} \tag{2}$$

The velocity of the particle is proportional to the element of accelerating voltage. As it results from the equation (2), the velocity value of the particle can be increased by increasing the field strength *E* or by increasing the distance *d* where the acceleration occurs.

Change of the particle trajectory

Electric field may influence the particles in motion by changing their trajectory of motion. Let's assume that electric field is generated by two plates and the voltage which is led to them is *U*; then *E=U/d*. May the particle of *m* mass possess a *q* charge and initial velocity of *v0* which is perpendicular to force lines. According to the principles of electrostatics, the particle is influenced then by force perpendicular to *v0* of *F=qE* value which will cause the curving of this trajectory according to the equation:

$$\mathrm{y} = \frac{qE}{2vm_0^2}\, x^2 \tag{3}$$

At the escape of the particle from the *x=l* field, and velocity in this place is expressed by the following equation:

$$v = \sqrt{v_0^2 + v_y^2} \tag{4}$$

Where v_y can be determined based on the equation (2) and then we obtain:

$$v_y = \sqrt{\frac{2qEd}{m}}\, y_0 = \frac{qUl}{mv_0 d} \tag{5}$$

The phenomena described above will constitute a base of the element used for data processing.

The architecture of Braun's machine

In this paragraph we will present a model of a processing element whose functioning uses a physical phenomenon described above. The functioning of Braun's machine is mainly modelled on the principle of functioning of Braun's cathode ray tube, so logical and physical construction of this processing clement is similar to the construction of the tube.

Element of Braun's machine

In figure Fig 1. a model of Braun's machine has been presented. Similarly to the tube, the model is equipped with an accelerating grid and a set of deflection plates and a visualizing space. In the described model, a particle gun is responsible for generation of particles. The gun is an equivalent for a cathode in the tube.

Accelerating plates

Plates are very important elements in the process of data processing. The task of accelerating plates is to accelerate particles which have been released by the gun. The value of the initial velocity of a particle is determined by the value of a control signal led to this plate. This signal corresponds to the accelerating voltage in Braun's cathode ray tube.

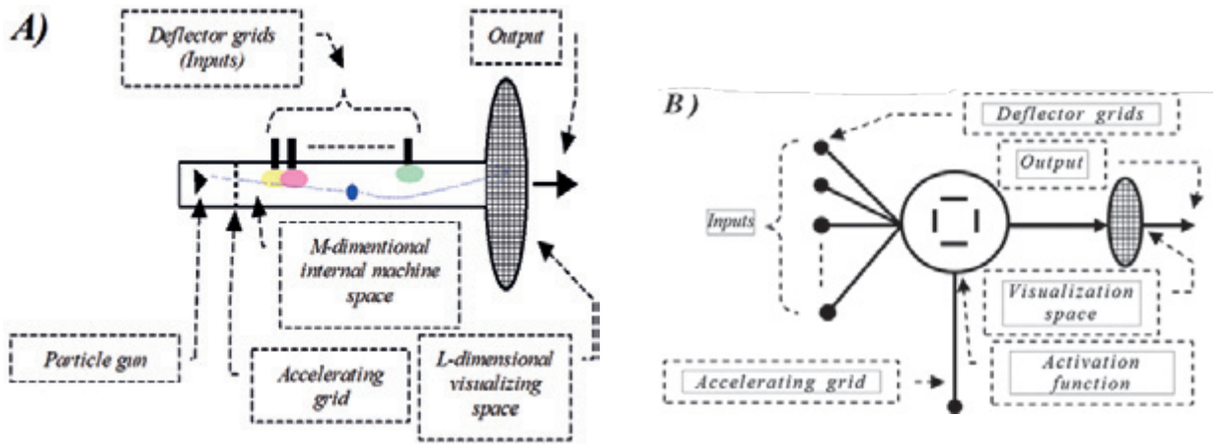


Fig. 1. A model of *M*-dimensional Braun's machine with *L*-dimensional visualizing space a) physical construction of a machine, b) logical construction of a machine.

Deflection plates

A set of deflection plates is the main processing element of the model being described here. The number of deflection plates corresponds to the number of inputs the machine has been equipped with. The size of the input signal led to the input of the machine corresponds to the value of deflecting voltage in a physical equivalent of the machine. The task of deflection plates is to change of the trajectory of a particle. The scope of changes in the trajectory of particles depends on the value of deflecting voltage led to deflection plates (5) which in the model we are describing corresponds to the values of input data.

Data processing mechanism

In the model being described in this paper data processing consists in subsequent modifications of the trajectory of a particle as a result of electric fields which influence the particle. The operation of deflection plates is the source of this influence. The value of electric field intensity is proportional to the value of input data. The higher values of data led to deflection plates we observe, the higher deflection it is. The result of data processing by a machine will be the value of final location of a particle in the space where it moves.

Considering the equations (5) it can be stated that the higher velocity of the particle, the smaller influence of electric field of deflection plates we can observe. Therefore, the influence of the values of input data on the processing operation and modification of the trajectory of a practice will be smaller. Increasing the velocity of a particle, we have an influence on generalisation of input data.

Visualizing surface

The last element of the presented Braun's machine model is the object which is responsible for visualization of the final location of a particle. This object is called a machine's visualizing surface. The functions of this clement are the same as the functions of a screen in Braun's cathode ray tube. It can be said that the screen of Braun's tube provides projecting operations of the location of a particle in the three-dimensional space on the plane.

A model of N-dimensional and M-input Braun's machine

After presenting a general idea of functioning of Braun's machine we will draw our attention to creation of a generalized model. Generalization of this model refers to a number of important aspects. The first aspect refers to the relation between the number of machine inputs and the number of deflection plates. It is a very important notion in case if N-dimensional data are to be processed by a machine. The second aspect is connected with the size of the internal machine space in which particles are moving. The size of this space decides on the classification qualities of a machine.

Construction of a function of machine activation

The description of machine operation presented in the previous paragraph implies that the process of data processing consists in repeated performing of three types of operations on a particle, namely, an operation of acceleration of deflection and visualisation.

$$X = (x_1 \times x_2 \times \ldots x_M) \tag{6}$$

$$V = (v_1 \times v_2 \times \ldots v_M) \tag{7}$$

Each of the operation mentioned here has an effect on the final location of a particle. In such a case a function of machine activation will be a creation of accelerating function, deflecting function and visualising function. Assuming that the accelerated particles move in *M*-dimensional space, then location vector *X* and velocity *V* of a given particle can be described using the equations (6, 7).

Accelerating function

In order to make the processing take place in a machine, a particle should be accelerated to a particular velocity. As it results from the aforementioned description, the value of this velocity decides on the course of the data processing operation performed by machine. The particle acceleration operation is described by means of accelerating function. Definitions of the accelerating function f_a were depicted using the two equations (8) and (9). The adopted notation of V^i indicates that the argument of a given function the i-th component of velocity vector *V*.

$$fa_j : (V_t^i, u_a) \rightarrow V_{t+1}^i \tag{8}$$

$$fa\,(V_t^i, u_a) := \sqrt{(V_t^i)^2 + K_a u_a} \tag{9}$$

The function (8, 9) provides changes of i-th coordinate of velocity vector *V*. As it results from the presented equation (9), velocity that will be gained by a particle because of its acceleration depends on two factors. First of all, it depends on velocity possessed by a particle before entering the area where the grid influences it. Secondly, it depends on the value of accelerating voltage. In the model we are describing, a numerical value is a counterpart of the value of accelerating voltage. In the equation (9) in comparison to the equation (3) there are no values of electric charge of a particle and mass. The reason for that is the fact that these quantities in the presented model have been adopted as constant values and are represented by means of a constant K_a .

Deflecting function

Another operation that is performed on a particle during the process of data processing is an operation which modifies the trajectory of a particle motion depending on the value of input data. In a machine, deflation plates carry out this operation. A number of modifications of the trajectory of a particle motion

corresponds to the number of deflation plates the machine is equipped with, but the number of plates depends on the dimension of input data and corresponds to the number of machine inputs. The equations (10, 11) define a function which performs the operation of deflection. A deflecting function *fd* possesses two arguments.

$$fd_j : (V_t, u_d) \rightarrow V^i_{t+1} \tag{10}$$

$$fd_j (V_t, u_d) : = \frac{K_d u_{d_j}}{\sqrt{v_1^2 + v_2^2 + \ldots v_M^2}} \tag{11}$$

The first argument is a velocity vector of a particle *V* which it has upon the moment of input in the area of influence of a given deflection plate. The second argument is the value of deflecting voltage u_d which corresponds to i-th element of the input vector. The operation of deflection described by *fd* function modifies the value of velocity vector *V*. During this operation, only this component of *V* vector is modified which belongs to the same plane as given deflection plates. Velocity for the selected coordinate of V vector is calculated as a quotient of deflecting voltage and total velocity of a particle. Similarly to the situation connected with a function which accelerates parameters which are constant, i.e. the size of electric charge of a particle, a mass of a particle and the width of deflection plates in formula (11) were not directly taken into account. The value of these parameters is represented by a K_d constant.

Visualization function

The last type of operation that is operated by a machine on a particle is the operation of particle visualization. Similarly to two previously discussed operations, this operation will also be described by means of a function. The operation of particle visualization is realized by means of *L*-dimensional visualizing surface. This surface is the counterpart of a screen covered with luminophore in Braun's cathode ray tube and, similarly to the screen, it is responsible for recording final deflections of a particle against the initial trajectory.

$$f_v : (V,(k_1 \times k_2 \times \ldots k_M) \rightarrow \overbrace{(x_{k_1} \times x_{k_2} \times \ldots x_{k_M})}^{L \subseteq \{1,\ldots M\}} \tag{12}$$

The formula (12) describes the operation realized by visualizing surface of a machine. Visualization function *fv,* playing the role of a screen, maps a set of values of M-dimensional velocity vector *V* of a particle on its final location *X* in a selected *L*-dimensional subspace which is the counterpart of L-dimensional screen.

Activation function

The activity of a machine consists in repeated performance of elementary operations on a particle in motion. Activation function describes the relation between input data and output data. For a Braun's machine we are describing in this paper, activation function is a combination of functions which correspond to elementary operations.

Let us use an example. Let us assume that we want to construct an activation function for a machine that is equipped with one accelerating grid, two deflection plates and with visualizing surface which is a plane. Apart from that, let us assume that the internal space of a machine, where a particle moves, is a three-dimensional space.

The aforementioned case, machine activation function is a compound of four functions, namely, an acceleration function, two deflecting functions and a visualization function. Because we assumed that the internal machine space is a three-dimensional space, machine deflection plates were located on two different planes. It should be, however, emphasized that the described machine model does not impose any established order connected with the location of plates in the internal machine space.

In a general case when a machine processes N-dimensional data, its activation function will be a compound of an accelerating function, N deflecting functions and a accelerating function. A machine space can have any size. The size of internal machine space has a decisive influence on separability of data processed. The increase of the size of internal machine space improves the quality of separation between classified data.

Algorithm for computing activation function

This algorithm computes the location of a particle in a visualizing surface. The location results from the influence of "electric fields" generated by input data of a machine on the particle.

1	Set the input parameters of the algorithm *M* – dimension of internal machine space; *L* – dimension of visualizing surface $V=(v_1,v_2,\ldots v_M)$-particle velocity vector; $X_{End}=(x_1,x_2,\ldots x_L)$ – location of a particle in a visualizing surface *QueueOfPlates* – set of deflection and accelerating plates of a machine *Input* – a set of input data *; i¬*0;
2	*Foreach plateÎ QueueOfPlates do*
3	*If plate* – is an accelerating grid *then*
4	u_a*¬U [plate];* // collect the value of accelerating voltage for *plate*
5	*axis¬SpaceOfMachine[net];*// collect the axis along which the *plate* operates
6	*V¬fa(V, axis ,* u_a*);* // Perform the operation of particle acceleration
7	*ElseIf plate* – is a deflection plate *then*
8	*i¬ i+1; //* subsequent number of machine input
9	u_d*¬Input[i];//* take the value of data for a given i-th input.
10	*axis¬SpaceOfMachine[plate];//* take the axis along which the *plate* operates
11	*V¬fd(V, axis ,* u_d*); //* Perform the deflection operation
12	End if; End foreach
13	X_{end}*¬fv(V,* $(i_1, i_2, \ldots i_L)$*);//* Perform the visualization operation where $i_1, i_2, \ldots i_L$ *Î{1,..M}*

Tab. 1. Algorithm for computing Braun's machine activation function.

In Fig. 2 we presented algorithm for computing the value of activation function. The operation of the algorithm consists in performing elementary operations on a particle that is moving in

the machine space. In order to do this, we check the set of plates a machine is equipped with. This set is organized in a form of a queue because it is important to remember that subsequent plates are collected in a specific order that corresponds to a physical location of accelerating and deflection plates inside the machine. After collecting a given plate from the queue of plates, the type of plate is checked. Depending on the type of the selected plate, specific operations for this type are performed. If the plate is an accelerating plate, we collect a value from a given machine input. This value is a counterpart of a deflecting voltage and is an argument of the deflecting function. In the next steps other attributes of a given plate are collected and the deflecting function is computed. When we perform all operations that were determined by the content of the *QueueOfPlates,* we perform visualization operation. This type of operation consists in mapping a velocity vector of a particle on its final location on the machine's visualizing surface. As it results from the presented algorithm, this surface is a subspace of the machine's internal space.

Braun's machine training algorithm

Machine training is a one-phase process. This process consists in computation of machine activation function for each vector of input data and in recording the location of a particle in the visualizing surface corresponding to this vector. In Fig. 3 we have presented Braun's machine training algorithm. This algorithm by the analogy to the phenomenon of luminophore burning in the cathode ray tube has been called a "burned surface" algorithm.

The training set is an input parameter of this algorithm. The set consists of vectors of input data together with a class number assigned to each vector. As it was already mentioned, a function of machine activation is computed for each data vector. The results of the computation of the activation function are the coordinates of a particle on the visualizing surface. A class number is assigned to each location of a particle on a surface and its environment where a given input vector belongs.

1	Set the parameters of training process – *TrainSet=ÈíInput x Classý*- train set – *BurnedSurface*- visualizing surface with the recorded location of particles
2	***For*** *i=1 to length(TrainSet)* ***do***
3	*(y_k….y_l)¬ fNB(QueueOfPlates, M, L, $TrainSet_i$®Input);// Compute activation function*
4	*BurnedSurface¬ BurnedSurfaceÈ{(y_k….y_l) ® Class };// Allocate class number to the location*
5	***End for***

Tab. 2. Braun's cathode machine training algorithm.

The algorithm enables generalization of input data in the process of training by means of decrease of influence of deflection plates of a machine as a result of increase of "accelerating voltage".

Response computing algorithm

The algorithm for computing responses was presented in Fig. 4. The following elements belong to its parameters: a set of machine's plates that describes its physical structure and burned surface with class numbers belonging to given points on this surface. This algorithm computes activation function for a given data vector. This function returns the coordinates of a particle and its environment on the basis of the physical structure of a machine and the value of input vector.

1	*(y_k….y_l)¬ fNB(QueueOfPlates, M, L, Input);// Compute activation function //(y_k….y_l) includes coordinates for an input vector on burned surface*
2	***If*** *BurnedSurface(y_k….y_l)¹ NULL* ***then***
3	Ans¬BurnedSurface(y_k….y);// return the class numbers belonging to the point and environment $_l$
4	***Else*** *//* failed attempt to classify input vector
5	Ans¬ NULL; // point (y_k….y_l) does not belong to any class
6	End if

Tab. 3. Algorithm for computing the response of Braun's cathode machine.

In the next step the algorithm maps the set of points on the class number which was assigned to this coordinates in the process of training. If the aforementioned mapping is not possible, algorithm signalizes a failed attempt of classification of input vector.

Testing Braun's cathode machine

In this paragraph we will present the results of the tests conducted on Braun's machine. These tests aimed at comparison of the presented machine model with well-known algorithms that are use with several basic notions connected with the classification of multidimensional data. The test task was divided into two categories. The first category is a classification where Braun's machine and selected comparative algorithm are trained to which class a given input vector should be classified during the process of training. During the training phase, algorithms classify data that belong to a testing set and undergo distortion process. All tests were conducted in MATLAB computational environment.

Classification

The first collection the Thyroid Disease is associated with medicine and the diagnosis of disease associated with thyroid. The other two data sets – the Landsat Satellite data and the Letter, they are related to image recognition. These data sets were taken from generally accessible data repository UCI Machine Learning Repository [1]. Detailed information about these data sets have been included [1, 3, 9]. Statistics relate to the selected data sets is shown in Table 1.

Tab. 4. Statistics of the test datasets [1, 9].

Data Set	Dimen-sionality	Number of classes	No of train samples	No of test samples
Thyroid	21	3	3772	3428
Landsat Satellite	36	6	4435	2000
Letters	16	26	15000	5000

Table 2 presents the results of the classification of these data sets using the Braun's machine and other algorithms. The results of other algorithms are derived from the work of [3, 9].

The table shows that, the effectiveness of the Braun's machine can be compared to the best algorithms presented in the table. The relatively bad results, the Braun's machine obtained for the Thyroid data set. In this case, it should be noted that the number of correct classification obtained using the machine, it is worse by only two percent of the best result. In contrast, the classification of other data sets, Braun's machine, was extremely effective in comparison with other algorithms. For the Letter data set, the machine has achieved the best result. In this case, the difference between the result obtained by the machine, and the third result, is more than two percent

The results in Table 2 show that the effectiveness of the machine is dependent on dimensional burned surface, value of accelerating voltage U_a and deflecting voltage U_d. In the case of the data set the letter, an increase of dimensional burned surface, has increased the efficiency of the machine by ten percent. The reason for such a phenomenon is the fact that the dimension of the burned surface increases separation between individual data classes.

Tab. 5. The percentage of correct classifications for the data sets. BM–Braun's machine with *L*–dimensional burned surface. (U_a/U_d – means a quotient of the value of accelerating voltage U_a and deflecting voltage U_d).

Thyroid Disease Data Set Algorithm %Test	Landsat Satellite Data set Algorithm %Test	Letter Recognition Data Set Algorithm %Test
CART tree 99.36	MLP, 36 nodes, +SVNT 91.3	**BM *L=14, U_a/U_d=1*** **96.06**
SSV tree 99.33	MLP, 36 nodes, 91.0	**BM *L=10, U_a/U_d=10*** **95.18**
MLP+SCG, 4 neurons 99.24	kNN, k=3, Manhattan 90.9	ALLOC80 93.60
SVM Minkovsky kernel 99.18	**BM *L=33, U_a/U_d =9*** **90.7**	K-NN 93.20
MLP+SCG, 4 neurons, 45 SV 98.92	FSMneurofuzzy, learn0.95 89.7	LVQ 92.10
FSM 10 rules 98.90	kNN, k=1, Euclidean 89.4	Quadisc 88.70
MLP+SCG, 12 neurons 98.83	SVM Gaussian kernel 88.4	CN2 88.50
Cascade correlation 98.5	RBF, Statlog result 87.9	Bayesian Tree 87.60
MLP+backprop 98.5	**BM *L=31, U_a/U_d =10*** **87.8**	NewId 87.20
SVM Gaussian kernel 98.4	MLP, Statlog result 86.1	IndCART 87.00
k-NN, k=1, 8 features 97.3	Bayesian Tree 85.3	C4.5 86.80
BM *L=19, U_a/U_d =16* **96.93**	C4.5 tree 85.0	**BM *L=8, U_a/U_d =10*** **86.56**
Naive Bayes 96.1	SSV tree 84.3	DIPOL92 82.40
SVM Gauss, C=1 s=0.1 94.7	Cascade 83.7	RBF 76.60
BP+conj. gradient 93.8	LDA Discrim 82.9	Logdisc 76.60
1-NN Manhattan, 93.8	Kohonen 82.1	Kohonen 74.80
SVM lin, C=1 93.3	Bayes 71.3	Backprop 67.30

Summary

In this article we have presented a model used for a classification of multidimensional data in a broader sense, called Braun's machine. The model is the result of research conducted by the author on the use of physical phenomena in the classification of data. Functioning of this machine has been based on principles which govern the particles in motion in electric field. The internal structure of the machine presented ion this paper has been based on the architecture of a cathode-ray tube – Braun's tube. In subsequent chapters based on the architecture of Braun's tube a general model has been presented whose generalization refers to any dimension of data and any dimension of visualising space. In this paper we have presented the functioning principle of the generalized machine model and discussed thoroughly the functions of individual machine elements in the process of data processing performed by machine. For a machine model described this way a Braun's machine training algorithm has been proposed as well as response computing algorithms.

In the final chapter we have presented the results of the machine tests for the notions connected with the classification and self-organization of multidimensional data. The results we have turned out to be very interesting, particularly when it terms of the notions connected with self-organization. During this type of tests it appeared that the responses generated by Braun's machine were equally good as comparative algorithms, or even in some case their quality was better. On the basis of the conducted research it can be said that the machine model is a very interesting computational model due to its limited computational complexity. The time of training of this machine was at least several dozen times shorter than it occurred with other algorithms with the same or lower level of errors in testing phase.

References

1. Blake C.L., Merz C.J.: UCI Repository of machine learning databases, http://www.ics.uci.edu/mlearn/MLRepository.html, University of California, Irvine, Dept. of Information and Computer Science, 1998-2003.
2. De Castro L.N., Von Zuben F.J.: An Evolutionary Immune Network for Data Clustering. *Proc. of the IEEE SBRN* 2000, pp. 84-89.
3. Duch W.: Support Vector Neural Training, *Lecture Notes in Computer Science* 2005, Vol. 3697: 67-72.
4. Griffiths D.J.: Introduction to Electrodynamics (3rd ed.), Prentice Hall 1998.
5. Peltonen J., Kaski S.: Discriminative Components of Data. *IEEE Trans. on Neural Networks* 2005, vol. 16, pp 68-83.
6. Święcicki M., Wajs W., Wais P.: An artificial immune algorithms apply to pre-processing signals, Springer-Verlag, 2004 – LNCS 3037, pp. 703-707.
7. Swiecicki M.: An algorithm based on the construction of Braun's cathode ray tube as a novel technique for data classification, LNCS Vol. 5864/2009, ICONIP 2009 – 16th International Conference on Neural Information Processing, A. Leung, M. Lee, J. H. Chan, Springer-Verlag, 2009.
8. Schlick T.: Molecular modeling and simulation: an interdisciplinary guide. Berlin: Springer 2002.
9. Statlog Datasets: comparison of results: http://www.is.umk.pl/projects/datasets-stat.html, Department of Informatics Nicolaus Copernicus University, 2009.
10. Tipler P.: Physics for Scientists and Engineers: Electricity, Magnetism, Light, and Elementary Modern Physics (5th ed.), W. H. Freeman, 2004.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 47-53

GRADING BREAST CANCER MALIGNANCY WITH NEURAL NETWORKS

Łukasz Jeleń[1], Artur Lipiński[2], Jerzy Detyna[1], Michał Jeleń[2]

[1] Institute of Materials Science and Applied Mechanics, Wrocław University of Technology, Smoluchowskiego 25, 50-370 Wrocław, Poland;
[2] Zakład Patomorfologii i Cytologii Klinicznej, Akademia Medyczna we Wrocławiu. Borowska 213, 50-556 Wrocław, Poland

Abstract: Breast cancer is one of the most often diagnosed cancers among middle-aged women. It is a well known fact that the early diagnosis is crucial and allows for the successful treatment while cancers diagnosed in their late stage are almost impossible to treat. For precise and objective diagnosis there is a need for a computerized method for malignancy grading, which is an integral part of a diagnosis process. In this work we present a classification system for grading cancer malignancy based on the Bloom – Richardson grading scheme. This is a well known grading scheme among the pathologist used during the diagnosis process. To achieve such a classification we extracted 16 features that were then used to classify the malignancy into two classes. Each class represents the malignancy of the cancer according to Bloom – Richardson grading scheme. According to that scheme two types of features are considered, where each type is extracted from images recorded at two different magnifications. Three structural features were calculated from low magnification images and thirteen polymorphic features were derived from high magnification images. To classify the malignancy grades, the multilayer perceptron was used. The described system was able to classify the malignancy with the error rate of 13.5%. In this paper we also present first clinical trials that allow for the verification of the obtained classification rate. The clinical trial showed that the depicted system has a high performance achieving an accuracy of 93.08% .
Keywords: Bloom – Richardson, malignancy grading, malignancy classification, cancer classification

Introduction

Breast cancer is the most often diagnosed cancer among middle-aged women. According to the World Health Organization there are 7.6 million deaths worldwide due to the cancer each year out of which 502,000 are caused by breast cancer alone. With such a high rate, breast cancer also is one of the most deadly cancers. Successful treatment is a key to reduce the high death rate. Most of the diagnosed cases can be fully recovered when diagnosed at an early stage. Cancers in their early stages are vulnerable to treatment while cancers in their most advanced stages are usually almost impossible to treat.

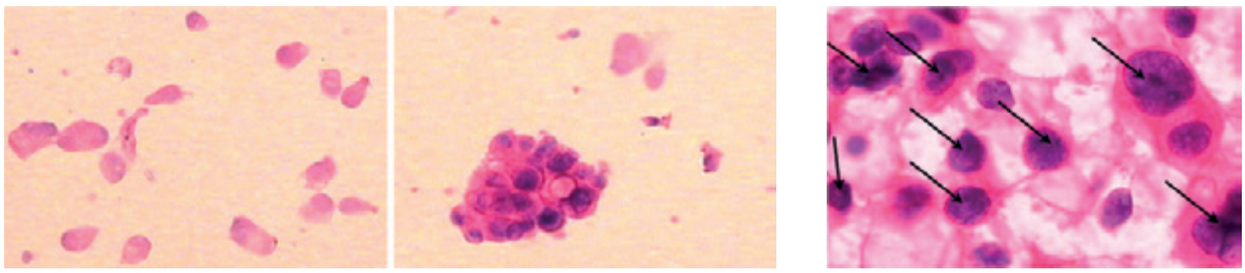
During the diagnosis process, the cancer is assigned a malignancy grade that is used to determine the appropriate treatment. Malignancy grading allows doctors to precisely estimate cancer behavior with or without undertaking treatment and therefore is called a prognostic factor. It plays an important role in breast cancer diagnosis and the appropriate treatment is chosen accordingly to this factor. This is a complicated process that involves assessing numerous nuclear features that allow for malignancy estimation. Cancer malignancy is graded based on a numeric scale that was introduced by Bloom and Richardson in 1957 [1]. The grading scheme proposed by the authors was derived to assess malignancy from histological slides and is now widely used among pathologists to grade not only histological but also cytological tissue.

The grading system originally proposed by Bloom and Richardson [1], later modified by Scarff and known as modified Scarff-Bloom-Richardson system, for grading breast cancer malignancy is one of the best known prognostic factors for this type of cancer [2]. These systems are based on grading of cells' polymorphy, ability to reform histoformative structures, and mitotic index. All of these features are described by the Bloom-Richardson scheme as three factors that use a point based scale for assessing each feature according to the following description:

1. **Degree of structural differentiation (SD)** – In histopathological slides this is also described as tubule formation, which reflects cell tendency to form tubules. Since in cytological smears

tubules are not preserved, the scoring given below for this factor is based on the classification of cell groupings within a smear, see Fig. 1a for example. On the right image in the Fig. 1a only one group is visible, which indicates lower malignancy than the case in the left image where dispersed cells are visible.

– One point cells in the image are grouped and spread regularly.
– Possibly instead of "grouped" it would be better to write "distributed". This is due to the fact, that cells "grouped regularly" can be recognized as "separated", as well.
– Two points – both grouped and single cells found within the image.
– Three points – cells are spread irregularly.



a) b)

Fig. 1. FNA smears. a) Structural differentiation. b) Pleomorphism.

2. **Pleomorphism (P)** – This factor takes into consideration differences in size, shape and staining of the nuclei. This scoring is fairly straightforward because with the growth of irregularity of the nuclei the prognosis becomes worse. Fig. 1b shows an example of these variations. Arrows in the image indicate cells with visible variations in shape and color.

– One point – nuclei with uniform size, shape and staining.
– Two points – moderate variations are found.
– Three points – very significant variations (see Fig. 1b).

3. **Frequency of hyperchromatic and mitotic figures (HMF)** – Mitosis is a process in the cell life cycle in which a mother cell divides into two identical cells. Main objective of this factor is to assess the number of mitosis in the field of view. Several fields of view on the same slide are taken into account because this step is done in a large magnification. The more cases of mitosis found, the worse the prognosis is. During the staining process, mitotic cells stain the most intensively providing the darkest areas in the nucleus.

– One point – occasional figures per field are found.
– Two points – smears with two or three figures in most fields.
– Three points – more than three figures per fields are found

According to the BR scheme, the malignancy of the tumor is assigned a grade that depends on the quantitative values of the above factors and is determined by the following equation:

$$G = SD + P + HMF. \quad (1)$$

The final grade is obtained by the summation of all the awarded points for each factor described earlier. Depending on the value of G, the tumor is assigned one of three grades according to the chart shown in Fig. 2.

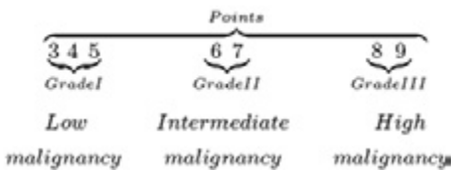


Fig. 2. Grade determination for the Bloom – Richardson scheme, taken from [2].

Based on the evaluation of the malignancy of the tumor an appropriate treatment is suggested.

Assigning a malignancy to a case is a very difficult task and is dependent on the experience of the pathologist. More experienced pathologists that have seen more cases are more reliable in their diagnosis. On the other hand, due to overwork and fatigue, seeing more similar cases may lead to misclassification of the malignancy. To address this problem we present an automated grading approach that is able to evaluate and assign a grade to Fine Needle aspiration biopsy (FNA) tissue. To achieve this we convert the Bloom – Richardson [1] grading scheme into a classification problem.

In the literature one can find approaches for breast cancer classification [3-9]. Most of these approaches involve classification of breast cancer as benign or malignant. In this work a system for classification of cancer malignancy is described. The proposed system implements the idea of multilayer perceptron to achieve the classification of malignancy into intermediate and high malignancy classes. To properly classify the malignancy a set of 16 features is extracted as described in section 2. In section 2.3 the reader can find a description of the classification procedure. The results obtained with the perceptron are then compared with those obtained by a pathologist as described in section 3.

Feature extraction and classification

Feature extraction is an important part of each classification task. Poor definition of features can lead to a high error rate of a classification system. Each classification system takes a feature vector as an input and responds with a category to which the object belongs. A feature vector is a set of features extracted from the input data. Before we can extract features used for classification our input data needs to be preprocessed and segmented. Preprocessing is a task of removing not important information from the data. Segmentation is an operation during which we isolate the boundaries of the important parts of the data that are then used for feature extraction and classification. In this study we make use of automatic thresholding as proposed by Riddler and Calvard [10] and fuzzy C–means (FCM) segmentation [11] to retrieve the nuclear information from fine needle aspiration biopsy (FNA) slides.

During the cytological examination two kinds of images are taken. The first type are the images that are recorded at 100x magnification. The second set of images is recorded at 400x magnification. All of the images used in this study had a resolution of 96 dots per inch (dpi) and a size of 764x572 pixels.

Due to the fact that different structures are visible at different magnification, The choice of the segmentation technique depends on the magnification of the image. For images recorded with low magnification the automated thresholding method shall be used (see Fig. 3a), while for the high magnification images the FCM segmentation will be applied (see Fig. 3b).

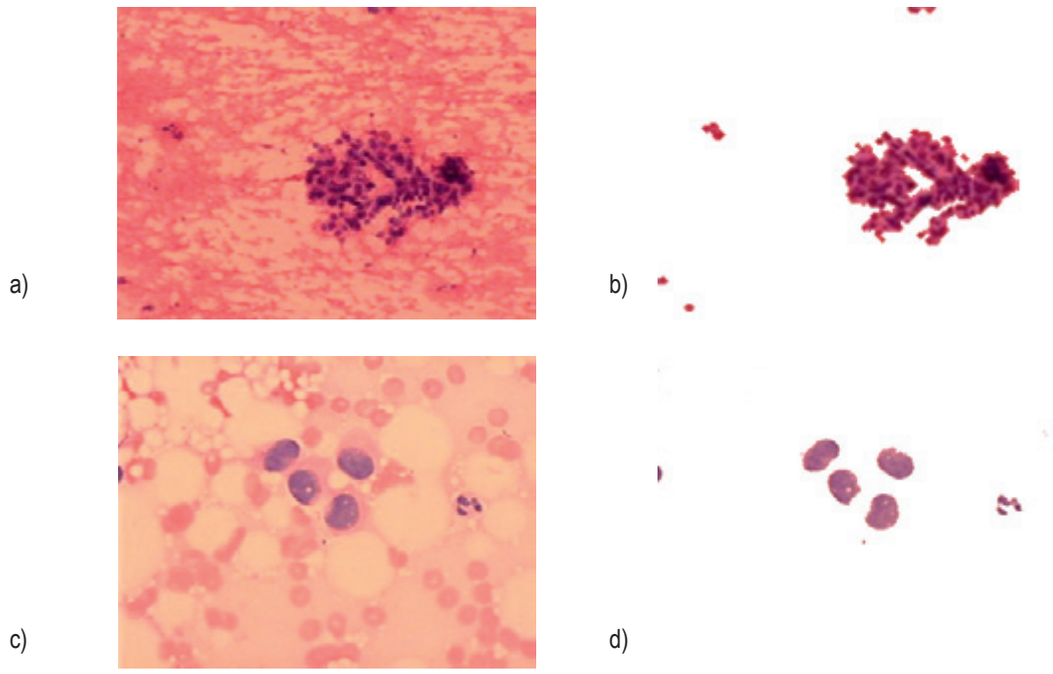


Fig. 3. Segmentation results. a) FNA slide at low magnification. b) Segmented image. c) FNA slide at high magnification. d) Segmented image.

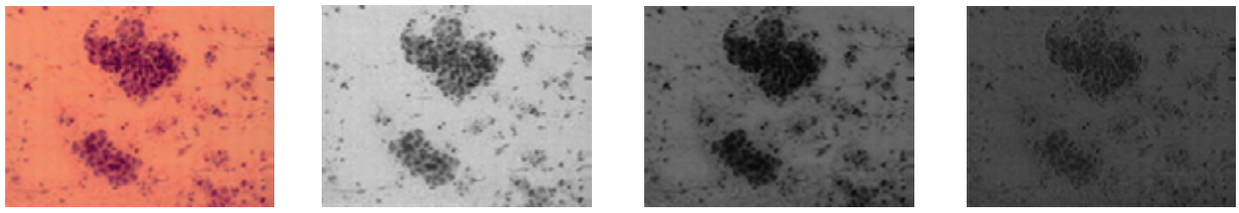
Based on the segmentation we are able to extract a set of 16 features that allow for training and testing the perceptron. For breast cancer malignancy classification two kinds of images are used. One subset consist of images recorded in low magnification, which allows for extracting structural features, the second subset of images is built from high magnification images which are used for polymorphic feature extraction.

Structural Features

Structural features are extracted from images recorded with 100x magnification. The information collected from these images allows us to extract three features based on cells ability to form groups or to be loosely spread around the image. Here, we segment those areas of the image where cells grouping are visible. For this purpose we make use of an iterative clustering approach for automatic image thresholding. This method was proposed by Riddler and Calvard [4]. In principle, their method seeks a threshold T, represented by a curve, within an image, that is restricted to have a bimodal histogram and the final threshold level is calculated according to the following equation:

$$T = \frac{\mu_1 + \mu_2}{2} \tag{2}$$

where μ_1 and μ_2 are the means of the components separated by T.



a) b) c) d)

Fig. 4. Comparison of RGB channels. a) Red channel. b) Green channel. c) Blue channel.

Due to the staining process of FNA images, the red channel provides best information about nuclear structures out of the three RGB channels (see Fig. 4). During the staining process, nuclei stain with shades purple and when red channel is extracted all the nuclear features are preserved while the background information is lost. This observation allows us to extract and threshold the image red channel and then to use it for further feature extraction.

For the purpose of this study we calculated three structural features. These features are defined, based on the number of groups and their area. Single cells, that are present in the images of high malignancy, are also represented as a group that consists of only one cell. Final decision is based on the relation between the features. If we take into consideration the description of Bloom and Richardson [1] of the dispersion measure and the nature of the images taken during the FNA examination (see Fig. 1) we can see that groups with larger area are less malignant that those with smaller areas. Analogically, we can say that images

with larger number of groups are more malignant that those that contain only one or few groups. Taking that into consideration we propose the following three features as a measure of cancer malignancy:

– **Average area (A_s)** – is calculated as the average number of nuclei pixels. This feature represents the tendency of cells to form groups. If As is large then there is one or few large groups in the image.
– **Number of groups (NG)** – To measure this feature we calculate the number of groups in the image that weren't removed during segmentation process. Here, if NG is large then there are numerous groups in the image, which suggests high malignancy case.
– **Dispersion (D)** – We define the dispersion as a variation of cluster areas
– (A_i) which is determined by the following equation:

$$D = \frac{1}{n-1}\sum_{i=1}^{n}(A_i - A_S)^2,$$

– where n is a number of cell clusters in the image and A_s is a mean area of clusters.
– Large values of this feature represent less disperse cells and therefore lower malignancy of the caner.

Polymorphic Features

To extract polymorphic features, high magnification images of breast cancer FNA slides were used. These features need more sophisticated segmentation method than structural features due to the fact that shape and staining information of nuclei is needed. To be able to extract shape–based features we make use of the automated segmentation procedure that involves the fuzzy approach of Klir and Yuan [11] that can be used to partition the image information to extract nuclei called the fuzzy c–means segmentation. In general, a set of data X = $\{x_1, x_2, \ldots, x_n\}$ is supposed to be divided into c clusters with assumption that P = $\{A_1, A_2, \ldots, A_c\}$ is known pseudo–partition and A_i is a vector of all memberships of x_k to cluster i. Now, using equation 3 the centers of the c clusters can be calculated [12].

$$v_i = \frac{\sum_{k=1}^{n}[A_i(x_k)]^m x_k}{\sum_{k=1}^{n}[A_i(x_k)]^m}, \quad i = 1, 2, \ldots, c$$

where $m > 1$ is a weight that controls the fuzzy membership. The memberships are defined by equation 4 if $||x_k - v_i||^2 > 0$ for all $i \in \{1, 2, \ldots, c\}$ and if $||x_k - v_i||^2 = 0$ for some $I \in I \subseteq \{1, 2, \ldots, c\}$ the memberships are defined as a nonnegative real number satisfying equation 5 for $i \in I$.

$$A_i(x_k) = \left[\sum_{j=1}^{c}\left(\frac{|x_k - v_i|^2}{|x_k - v_j|^2}\right)^{\frac{1}{m-1}}\right]^{-1},$$

$$\sum_{i \in I} A_i(x_k) = 1.$$

The clustering algorithm seeks for a P that minimizes the performance index $J_m(P)$ which is defined by equation 6 and the optimization solution to this problem can be found in [13].

$$J_m(P) = \sum_{k=1}^{n}\sum_{i=1}^{c}[A_i(x_k)]^m |x_k - v_i|^2.$$

The presented segmentation technique can be applied to a color image without additional preprocessing techniques. Based on the segmentation results obtained with the described method 13 polymorphic features were extracted:

– **Area (A)** – calculated as the sum of all pixels of the nucleus (N_i) [14],
– **Perimeter (p)** – the length of the nuclear envelope. Calculated as a length of the polygonal approximation of the boundary [14].
– **Convexity (C)** – calculated as the ratio of nucleus area and its convex hull [14], see Eq. 7

$$C_i = \frac{A_i}{Area(H(N_i))}.$$

– **Orientation (Or)** – provides us with the information about the orientation of the nuclei. It is considered as an axis of least inertia. When the coordinate system is placed at (x_i, y_i) then the orientation Or_i can be defined as

$$\Theta_i = \tan(2\theta_i)$$

– **Vertical projection (v_i)** – this feature is calculated as a sum of all pixels along columns of the nucleus image [14]. summation of all the columns determines the vertical projection.
– **x – centroid** – finds a center point of a nucleus along each row. When combined with the y – centroid, it is often called a center mass of the object.
– **φ3** – this is feature calculated based on the image moments and is also called a momentum feature. Moments are generally used for the extraction of features that are rotation, scaling and translation (RST) invariant. Based on the normalized central moments, η_{ij} , this feature can be calculated as:

$$\varphi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2.$$

– **Histogram mean, histogram energy** – these two features are calculated based on the histogram of the image, which provides statistical information about the gray levels distribution within an analyzed image [14].
– **Textural homogeneity** – is a statistical feature that is calculated from the textural information of the image. The determination of the feature depends on the gray level co-occurrence matrix that poses the statistical information about the texture of the image [14].
– **Red histogram mean, red histogram skew, red histogram width** – this set of features is calculated from the image histogram. The histogram was calculated for the image red channel that provided the most valuable information about the nuclear structures [14].

Classification

Based on the features described in sections 2.1 and 2.3 a neural network was used for classification. The idea of neural networks is based on the real interactions of human nerve system. The basic element of the neural network is the neuron, or sometimes also called a perceptron. It is a mathematical model of a biological neuron [15]. Combining a few neurons together in such a way that the neurons can interact with one another make a neural network that is able to process input data and provide us with a certain decision.

Each neuron accepts an input signal of the form $X = [x_1, x_2, \ldots, x_n]$ and each of the sub-signals are assigned a weight. $F(s_i)$ is called an activation function of the neuron and depending on the type of the neuron activates its output. For a perceptron, which is the simplest neuron model, the activation function is of the form:

$$y_i(s_i) = \begin{cases} 1 & f \quad s_i > 0 \\ 0 & f \quad s_i \leq 0 \end{cases}$$

where si is an output signal calculated from equation 11,

$$s_i = \sum_{j=0}^{N} w_j x_j$$

Another type of neuron is called a sigmoidal and it is characterized by an activation function of the form:

$$f(x) = \frac{1}{1 + e^{\beta x}}$$

Before we can use our neural network it is necessary to train it such that it will be able to recognize patterns. Training of a neural network is based on adjustments of weights depending on the output value. Multilayer perceptrons (MLP) are simple and one of the most widely used neural networks trained with a backpropagation method in a supervised manner [14]. These networks are powerful and are able to approximate arbitrary functions [16]. The backpropagation learning allows for error propagation through the network and adaptation of the weights of the hidden neurons. This error correction training assumes that the desired network response is known a priori, which is usually the case in pattern recognition.

In this study we trained the MLP with the feature vector built from the described features. This 16–element feature vector was presented at the input layer of the network and fed through one hidden layer to a 2-element output layer. For the training purposes 54 cancerous cases were used.

Results

This section shows the obtained results of breast cancer malignancy classification. In this study we have tested the MLP with cross validation method that is a statistical method of partitioning of the data into subsets for further analysis. In principal, one subset is used for testing while the remaining subsets are used for training the classifiers. In general, this is called a K–fold cross–validation. The idea behind K–fold cross–validation is that the data is partitioned into K subsets from which one is retained for testing and the remaining K–1 subsets are used for training. This process is repeated K times and each time a different fold is used for testing. For the purpose of this study a K=10 was used and the results of the validation are summarized in Table 1.

Test 1	17.01%	Test 6	5.84%
Test 2	41.65%	Test 7	2.44%
Test 3	4.78%	Test 8	3.33%
Test 4	13.11%	Test 9	20.37%
Test 5	6.32%	Test 10	19.48%
		Average	13.48%

Tab. 1. Error rates obtained with Cross-Validation.

Cross-Validation shows how well the data can be classified depending on a set that has been used for training. As we can see from the Table 1, the overall classification error is 13.48%. Among the ten training sets one can notice that the best classification was achieved for Test nr 7, which consist of the most distinctive feature values for both intermediate and high malignancy classes.

Additionally, the depicted system has been successfully introduced in the Pathological laboratory at the Medical University of Wrocław to assist doctors in their diagnosis. First results obtained from the clinical test show that the proposed system can classify the malignancy with a high accuracy. Table 2, shows the classification results obtained by the multilayer perceptron on the real data.

		Predicted	
		G2	G3
Actual	G2	148	11
	G3	24	11

Tab. 2. Confusion matrix.

From the Table it can be noticed that most of the intermediate cases (G2) were correctly classified. The achieved overall classification accuracy for intermediate cases was 93.08% while for high malignancy alone the accuracy was much lower. The average performance of the proposed system was 81.96%.

Conclusions

In this study a breast cancer malignancy classification system was described. The depicted system is able to classify malignancy into one of the two malignancy grades: intermediate and high malignancy. The low malignancy cases are very rare and therefore were not used in this study. From the results shown in section 3, it can be noticed that the multilayer perceptron is a good choice for the classification of breast cancer malignancy, what is also supported by the test performed during the clinical trials on a real data.

From the Table 1 we can see that the average error of the MLP achieved with cross-validation was recorded at a level of 13.48% which is a relatively small error rate for this kind of images. Fine needle aspiration biopsy images are difficult to grade for a pathologist and therefore any tool supporting the diagnosis is helpful. To verify the clinical applicability of the proposed system it was tested by a pathologist on a real data. According to that verification, the performance of the system is promising. The achieved error was 18.04%, which is just 4.56% higher that the results obtained with cross-validation. One can draw a conclusion that the system performance is marginally lower on the data that was not used on for training.

The obtained results show that the computer aided cancer malignancy grading can be used to help a pathologists during the cytological examination of the breast cancer. Looking at the nature of the images, the lower performance could be expected. From Table 2 one can see that there was 24 false positive classifications. Considering this fact we can conclude that either the G3 cases incorrectly classified where border cases, or the training set did not have a sufficient number of such cases. In the future research it would be necessary to include these difficult cases to the training set.

Over all we can conclude that from the pathologist point of view, having a system that is able to assist them during the decision making process, is very useful and helpful especially in the situations where it is difficult to decide to which malignancy class the case should belong. Such a system would help to make their diagnosis more objective and precise.

References

1. Bloom H.J.G., Richardson W.W.: Histological grading and prognosis in breast cancer. *British Journal of Cancer* 1957, 11: 359-377.
2. Le Doussal V., Tubiana-Hulin M., Friedman S., et al.: Prognostic value of histologic grade nuclear components of Scarff–Bloom–Richardson (sbr). An improved score modification based on a multivariate analysis of 1262 invasive ductal breast carcinomas. *Cancer* 1989, 64(9): 1914-1921.
3. Street N.W.: Xcyt: a system for remote cytological diagnosis and prognosis of breast cancer. In: L.C. Jain (ed.), Soft Computing Techniques in Breast Cancer Prognosis and Diagnosis, pp. 297-322, Singapore 2000. World Scientific Publishing.
4. Street W.N., Wolberg W.H., Mangasarian O.L.: Nuclear feature extraction for breast tumor diagnosis. In: Imaging Science and Technology/Society of Photographic Instrumentation Engineers 1993 International Symposium on Electronic Imaging: Science and Technology, IS&T/SPIE, Vol. 1905, pp. 861-870, San Jose, California 1993.
5. Nezafat R., Tabesh A., Akhavan S., Lucas C., Zia M.A.: Feature selection and classification for diagnosing breast cancer. Proceedings of International Association of Science and Technology for Development International Conference, IASTED, Cancun, Mexico 1998, pp. 310-313.
6. Walker H.L., Jr., Albertelli L.E.: Breast cancer screening using evolved neural networks. IEEE International Conference on Systems, Man, and Cybernetics, San Diego 1998, USA, 2: 1619-1624.
7. Cheng H.D., Li X.Q., Riodan D., Scrimger J.N.: A Parallel Approach to Tubule Grading in Breast Cancer Lesions and its VLSI Implementation. Computer-Based Medical Systems: Fourth Annual IEEE Symposium 1991, pp. 322-329.
8. Schnorrenberg F., Tsapatsoulis N., Pattichis C.S., Schizas C.N., Kollias S., Vassiliou M., Adamou A., Kyriacou K.: A modular neural network system for the analysis of nuclei in histopathological sections. *IEEE Engineering in Medicine and Biology Magazine* 2000, 19: 48-63.
9. Jeleń Ł., Fevens T., Krzyżak A.: Classification of Breast Cancer Malignancy using Cytological Images of Fine Needle Aspiration Biopsies. *Int. J. Appl. Math. Comput. Sci.* 2008, Vol. 18, No. 1: 75-83.
10. Ridler T., Calvard S.: Picture thresholding using an iterative selection. *IEEE Trans. System. Man and Cybernetics* 1978, 8: 630-632.
11. Klir G.J., Yuan B.: Fuzzy sets and fuzzy logic: Theory and applications. Prentice Hall 1995, New Jersey, .
12. Theera-Umpon N.: Patch–Based white blood cell nucleus segmentation using fuzzy clustering. *ECTI Transactions on electrical eng., electronics and communications* 2005, 3(1): 15-19.
13. Bezdek J.C.: Pattern recognition with fuzzy objective function algorithms. Plenum Press: New York 1981.

14. Jeleń Ł.: Computerized cancer malignancy grading of fine needle aspirates. PhD thesis, Concordia University 2009.
15. Rangayyan R.M.: Biomedical Image Analysis (Biomedical Engineering). CRC Press: Boca Raton, FL 2004.
16. Lapedes A., Farber R.: Nonlinear signal processing using neural networks: prediction, and system modelling. Technical Report, LA–UR–87–2662, 1987.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 55-66

COMPARING APPLICATION OF THE WAVELET TRANSFORM AND THE ADAPTIVE WEIGHT SMOOTHING ALGORITHM FOR THE NOISE REDUCTION IN MAGNETIC RESONANCE IMAGING

Joanna Świebocka-Więk, Henryk Figiel

AGH–University of Science and Technology, Faculty of Physics and Applied Computer Science, Al. Mickiewicza 30, 30–059 Cracow, Poland

Abstract: The analysis of the Adaptive Weight Smoothing (AWS) and the Discrete Wavelet Transform (DWT) application in the MRI images improvement is presented. The results show how the AWS and the DWT algorithms can be used for the noise reduction in the MRI. The DWT application gives much better visual results without blurring. It is shown that for the DWT a bigger number of the input parameters (like the wavelet's choice, number of the iteration, degree of the decomposition, kind of the smoothing thresholding etc.) is needed, what makes it more difficult to optimize the output image. The analysis of the Signal to Noise Ratio (SNR) and the Contrast to Noise Ratio (CNR) and edge quality detection for both methods is presented to show their effectiveness. We concluded that the AWS algorithm can be applied to improve MRI images mainly in the case of high noise and low Signal to Noise Ratio (SNR) whereas wavelet transform is effective in any case. The wavelet transform application provides additional possibilities like image compression and image fusion, which can be useful in the MRI.
Keywords: magnetic resonance imaging, noise reduction, edge detection, adaptive smoothing, Wavelet Transform

Introduction

Finding the best way of the Signal to Noise Ratio (SNR) improvement in the Magnetic Resonance Imaging (MRI) is a key element of the image quality correction. Among the broad spectrum of methods, the numerical methods are very useful.

Recently the Adaptive Weight Smoothing algorithm (AWS) was proposed by J. Polzehl and V. Spokoiny [1, 2]. It has been already tested for the radar images because of their very characteristic noising known as the *speckle effect* [3]. Although this method is dedicated to the images with the Gaussian data distribution it can be applied in case of the MRI images with the Rician data distribution which is signal–dependent; the image areas with low intensities can be represented by Rayleigh distribution and the image areas with high intensities can be represented by Gaussian distribution as well. Additionally, Rician distribution can be approximated by the Gaussian distribution depending on the image SNR (generally for SNR greater than 15 dB). In addition it is worth to examine AWS application to improve MRI images, because it has not been done so far.

Other, alternative method is using the wavelet transformation, based on the decomposition of the image applying high– and low–pass frequency filtration. Earlier publications about this method application in the MRI reported encouraging results [4], therefore a further improvement is worth of interest.

In our analysis both methods were implemented in the MATLAB environment. The main goal of this publication is to show, compare and rate the results of these methods in MRI, taking into account the chosen image quality criteria like: visual evaluation, Signal to Noise Ratio (SNR) and the Contrast to Noise Ratio (CNR), degree of keeping the edges and typical structures, degree of smoothing the image. Research was performed using bitmaps, fantom images and diagnostic images.

The Adaptive Weight Smoothing algorithm

Adaptive Weight Smoothing algorithm requires two input parameters: l which determines the power of image smoothing in each iteration step and hmax which defines the maximum pixel surrounding. This method is local; in the each iteration step for every pixel, intensities of surrounding pixels are modified by the sets of weights. This method is also propagational; in each next iteration step, h value is increased (wider surrounding) and the pixel surrounding range can be modified as well. The set of weights depends on the difference in the intensities of pixels in

previous steps of the iteration and it can be controlled by the parameter λ. Best results are for λ= 2 – 4. The number of the sets of weights is equal to the number of pixels in the image. For each pixel the number of the weights in each set is equal to the number of pixels in its surrounding.

Algorithm description

AWS algorithm has four basic steps:

1. Initialization

For each pixel a set of weights is calculated. The number of such sets is equal to the number of pixels in the image and the number of weights in one set is equal to the number of pixels in the surrounding. At this point the first set of weights is calculated. In the end (and also after every iteration) a new radius is calculated.

2. Weights adaptation

In each iteration (except the first one), for each pixel in the surrounding, there are calculated three parameters: weight (basing on results from the previous iteration), local transition (defined by the Euclidean distance between a pixel and a pixel in its surrounding) and the statistic transition (defined by λ parameter and an intensity estimator function for the previous iteration).

3. Local estimation

In each iteration (except the first one) the intensity estimator is calculated. Radius of the surrounding increases.

4. Termination

It takes place for $h_{(k)} >= h_{max}$, where k is the number of the iteration. The maximum radius h_{max} affects the number of all iterations

Figure 1 ilustrates the pixel surrounding definition and the way of pixels numeration in the rows and in the columns. The first surrounding has a square shape and it consists of 8 pixels. In each next step the shape is preserved but the number of the pixels increases nonlinearly.

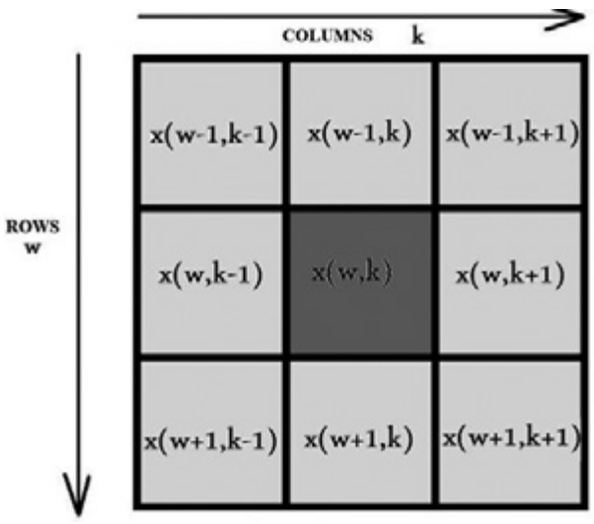


Fig. 1. Pixel surrounding definition.

Although AWS algorithm is effective, it is important to mention about some challenges in its proper implementation. The basic problem is a high, nonlinear time–complexity. Because the surrounding radius (the number of pixels belonging to this surrounding) increases nonlinearly, the calculation time increases in non linear way as well. It is also very important to find the best weight function because of its tremendous influence on the image smoothing. It is necessary to mention about the case where the pixel intensity does not change in subsequent iterations. In such situation, according to the basic algorithm version weight functions are clearing the pixel intensity, turning it into black. That is why some changes in algorithm were necessary to implement. There is also an interesting possibility of defining new surrounding types. For the purposes of this publication the basic surrounding contains 8 pixels and it has a square shape.

Merging the images

Merging the images is a very effective method of the image quality improvement. It is based on weighted fusion of three images according to the equation .

$$\delta = \frac{g_\alpha \alpha + g_\beta \beta + g_\gamma \gamma}{g_\alpha + g_\beta + g_\gamma}$$

where:

δ – output image
α –input image with the high noise level (full information about texture and edges)
β – image processed with the low I value (edges)
γ – image processed with the high I value (input image approximation, strong smoothing)
g_α, g_β, g_γ – weights of the individual images

The first image α is an input image without any noise reduction. The second image, named β, is the image after AWS filtration with the low λ value (the image is not much smoothed, the noise is only partly reduced, and the most important edges are kept). The last image γ is also the image after filtration, but with the high λ value (image is smoothed and partly blurred; noise is strongly reduced, only the sharpest edges are kept and visible). Weights coefficients allow modifying the contribution of the individual images in the final image.

The Wavelet Transform

Wave and wavelet definition

Fundmanetal difference between the Fourier Transform and the Wavelet Transform is the kernel of the transformation multiplied by the signal value. The Fourier Transform as the kernel of the transformation uses sine/cosine function whereas Wavelet Transform uses wavelets [6]. The main task for both methods is the approximation of any continuous function with the specified precision; in case of wavelets this approximation is expressed by so called *wavelets coefficients*.

Fourier Transform is used to frequency analysis of stationary (periodic) time series and to extract signal's global characteristic. However changing reference system from time–value to frequency–value, event's time information is being lost. That is why Fourier Transform is helpful in case of long lasting time series, when transient time location is irrelevant.

Wavelets are continuous oscillatory waveforms of different durations. Wavelets Transform is based on the signal decomposition by the wavelets created by scaling and shifting of the *mother wavelet*. Wavelet Transform serves to the time–frequency analysis of nonstationary time series and to extract global and local signal's characteristic. Changing reference system from time–value to time–scale (frequency) allows for the analysis of the frequency evolution in time domain. Although it is achievable to get exact information about the time of frequency changing, it is not possible to get information about specific frequency in signal, only about some frequency range. Wavelet Transform should be used in case of nonstationary signal, full of transients with specified frequencies for which time location is important. The diffrence between wave and wavelet function is showed in the Figure 2.

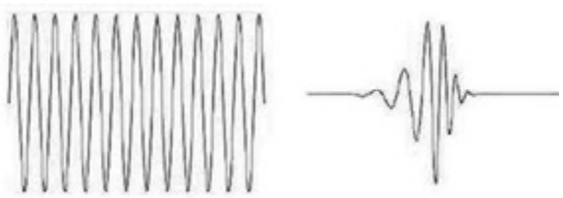


Fig. 2. Time course of the wave and the wavelet function.

Fundamentals of the Wavelet Transform

In the wavelet analysis (called also *the multiresolution analysis*) a set of two functions is needed: the wavelet function (responsible for high pass filtration and keeping the signal details) and assigned to it the scaling function (responsible for low pass filtration, signal smoothing and averaging). In each transformation only one wavelet function can be used. It is not possible to use several wavelets functions in the same time.

The wavelet function has also some limitations like zero average value:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0$$

Wavelet function should be also orthogonal and normalized:

$$|| \psi ||= 1$$

It should also have a finite bandwidth (duration) and its values at infinity must tend to 0.

Continuous Wavelet Transform (CWT)

Wavelet Transformation for continuous signals is the integral transformation. Calculated coefficients are the scalar result of the signal and the wavelet function, which allows determining how precise used wavelet function reflects the signal.

$$s_{\psi}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t)\psi(\tfrac{t-b}{a})dt$$

where:

s_{ψ}(a,b) – wavelet coefficient depending on *a* and *b* parameter (the scalar result of the signal **s(t)** and the wavelet, which allows to determine how well the wavelet approximate the signal,

a – a scale parameter (changes wavelet's duration),

b – a shifting parameter (changes the wavelet's location) ,

s(t) – time–depending signal,

Ψ(t) – a wavelet function.

As a result of changing *a* and *b* parameter value the wavelet family is created, which is used to decompose the signal *s(t)* on several detail's levels. Within the family all the wavelets have the same shape, accuracy of the scale and shifting.

The Continuous Wavelet Transform algorithm has two basic steps:

1. Chosen wavelet function is compared with the beginning part of the analyzed signal. Calculated coefficient determines the similarity between the wavelet and the current part of the signal. Then the next part of the signal is selected (increasing *b*) and compared with the wavelet. This process is repeated until the whole signal is analyzed.
2. Next wavelet is scaled (increasing *a*) and 1st step is repeated. The last action is taken when the duration of the wavelet is equal to the duration of the signal. To visualize the coefficients, the reference system time–scale (frequency) is used. The brightness of each point is proportional to the wavelet coefficient value.

Discrete Wavelet Transform (DWT)

The Continuous Wavelet Transform is very difficult to implement, especially for the discrete signals. For example it contains a lot of redundant information, particularly for the signal reconstruction. There are some methods of its acceleration; one of them is the Discrete Wavelet Transform (DWT). Signal is processed in the series of the cascade–connected filters. High–pass and low–pass filtration are used in the same time. Signal is decomposed into two subsignals representing low and high frequencies. Subsignal which contains low frequencies represents the input signal's approximation (basic informations about its structure). Subsignal which contains high frequencies represents image details (like edges or microstructures). Only subsignal with high frequencies is sampled again for new approximation and detail subsignal of the higher level. In each sampling step detail coefficients are expressed by the Wavelet Transform coefficients, and the approximation coefficients are expressed by scaling function coefficients.

Multiresolution analysis

The signal is represented as a sum of the approximation subsignal and subsignal representing details what generate increased number of received samples. It allows removing every second sample from the signal (Kotielnikov Shannon theorem) and receiving two-times decreased subsignals. This process, called *subsampling,* increases the scale by two times, but decreases resolution by two times as well. In the same time signal is passed through the high–pass filter, the received subsignal is after subsampling describes the input signal details. On each level, the previous level representation is again represented as a sum of the signal details and approximation. This action creates an array of the following coefficients called the *wavelet coefficients*. The number of these coefficients is equal to the number of the samples but their arrangement in the signal is specific – associated with the process of filtering and subsampling. Multiresolution analysis is based on a successive signal decomposition into the series of subsignals with a decreased resolution, on many levels.

In Figure 3 a multiresolution analysis scheme is presented.
s_l – approximation (low frequency signal representation),
d_l – details (high frequency signal representation),
h[n] – high- pass filter,
g[n] – low- pass filter
↓2 – subsampling.

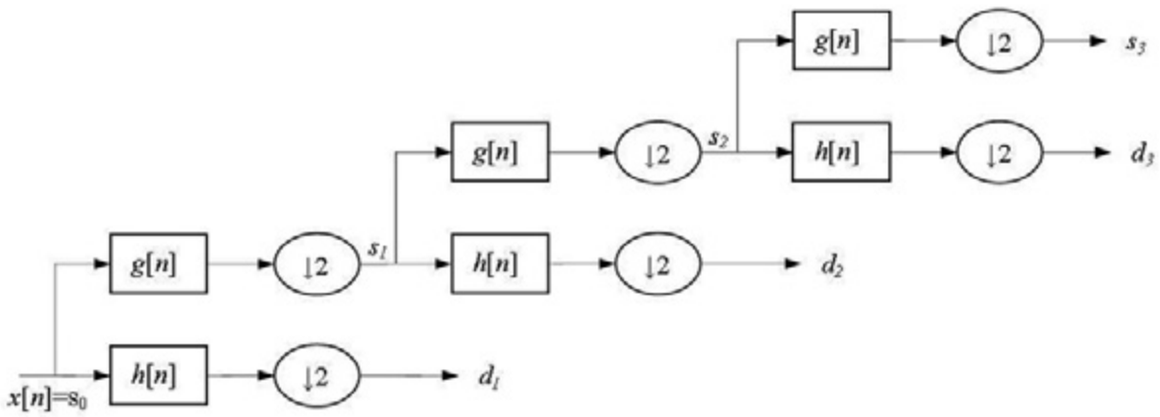


Fig. 3. Multiresolution analysis scheme. The input signal is divided into two subsignals s_l and d_l, each contains double–less samples, what allows a lossless signal reconstruction. Signal s_l is decomposed once again, although signal d_l is saved in the output of the transformation.

Mallat's algorithm

Mallat's algorithm is a method of calculating the scaling function coefficients and the wavelet function coefficients in the each step of the signal decomposition, basing on the well known functions from the previous step. It is often implemented for the multiresolution analysis in a two–dimensional space, like images. This method decomposes an input image into four component subimages. The decomposition is based on a sequential low–pass and high–pass filtering separately along the rows and columns (the rows are decomposed at first, next basing on received subimages columns decomposition is proceed). Image is represented by four matrices of the wavelet coefficients having twice less linear resolution. Each of these subimages can be decomposed again in the same way. This method allows for an implementation of the multi–level resolution and receiving the image representation on many detailed–levels.
Figure 4 presents the way of using the Mallat's algorithm for the image decomposition. LL means a low–pass filtering of the rows and columns (approximation). LH means a low–pass filtering of the rows, high–pass filtering for columns (keeping the vertical edges). HL means a high–pass filtering of the rows, low–pass filtering of the columns (keeping the horizontal edges). HH – high–pass filtering of the rows and columns (keeping the diagonal edges).

Results

For the purpose of our analysis the brain images were noised by using the MATLAB script functions, fantom images were naturally noised by image acquisition terms.

All used images were normalized before their processing and after each processing step. It is a process of adjusting the range of the input image data variation to specified range (in the case of the gray, 8-bit images it is from 0 to 255). It was done according to the equation:

$$g_{n,m} = g_{\max}\left(\frac{f_{n,m} - f_{\min}}{f_{\max} - f_{\min}}\right)$$

where:
$g_{n,m}$- the output pixel intensity,
g_{max}- the maximum output pixel intensity,
$f_{n,m}$- the input pixel intensity,
f_{min}- the minimum pixel intensity in the image,
f_{max}-the maximum pixel intensity in the image,
The basic function in MATLAB allowing the image normalization is the *imadjust* function.

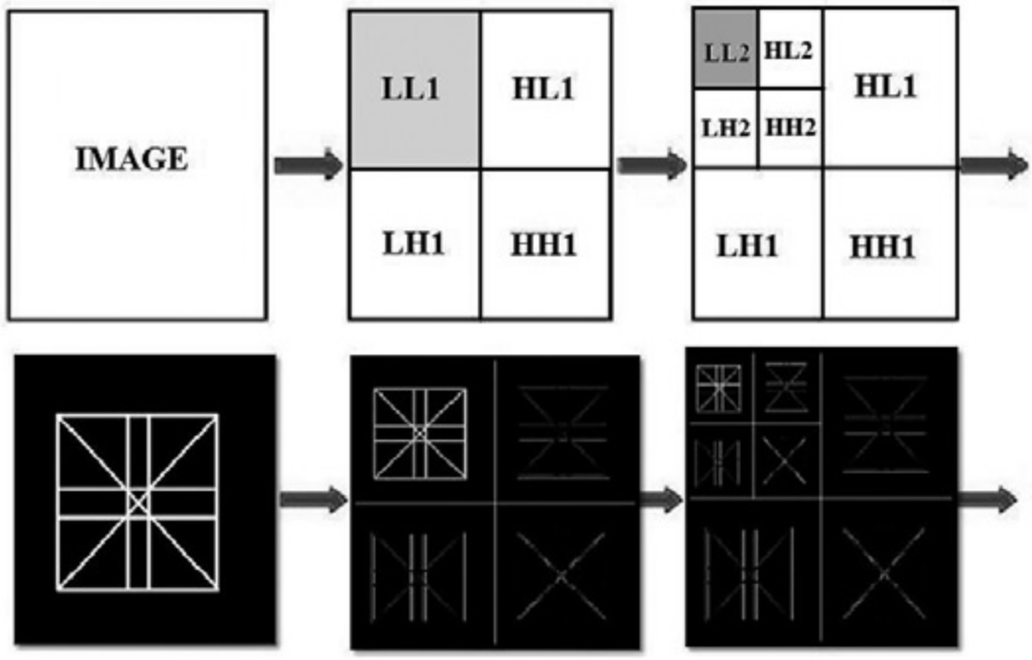


Fig. 4. Image decomposition scheme.

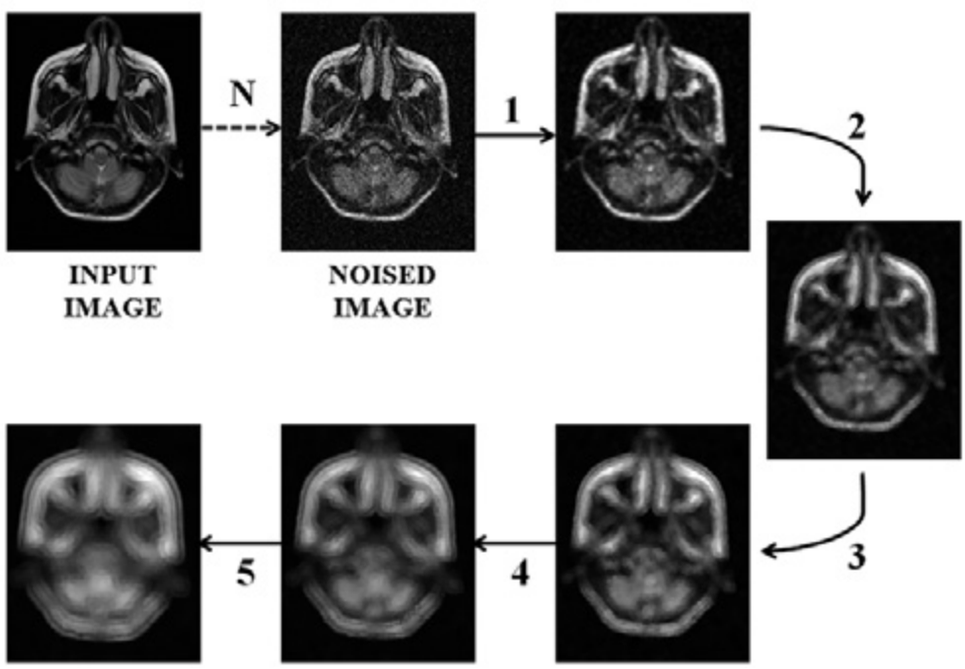


Fig. 5. The AWS algorithm functioning in the following iterations (λ=4, $h_{(k)}=a*h_{(k-1),}$ where k means number of the iteration, a =sqrt(1.25), each h was rounded to an integer).

Application of the AWS algorithm

Based on a literature concept of the AWS algorithm, detailed version, specially dedicated to the MRI application was written. The Figure 5 ilustrates the AWS algorithm functioning. In the beginning for a chosen image for the noise removing, the *λ* and h_{max} parameter values are defined. After the first iteration it is possible to see a visual effect; the image is smoothed and the noise is partly reduced. As it is shown in the Figure 5, in each iteration smoothing and noise reduction effects are stronger. Basing only on the visual evaluation it is advisable to terminate the algorithm after the 3rd or the 4th iteration. It is advisable also because of the high time–complexity of the algorithm; executing even a little iteration for a simple image takes even an hour. Finding the best, but not to high l value is needed to prevent too strong image smoothing which makes it useless from the medical

point of view. On the other hand if the I value is too small, the noise will not be reduced satisfactorily. That's why it is important to find a compromise between these two needs.

Application of the DWT algorithm

To proceed a noise reduction using the Wavelet Transform filtration, the Matlab Wavelet Toolbox was used, because of its huge number of possibilities like the image compression, image fusion or filtering signals in one and two dimensions. Noise reduction using the DWT method has three basic steps:

1. choosing the wavelet family and image decomposition on the chosen degree of decomposition (*N*),
2. thresholding discrimination of the details coefficients for all (from *1* to *N*) degrees of decomposition. For each degree (each frequency bandwidth) coefficients with low amplitudes (below the threshold) are removed from the image.
3. image reconstruction (using original approximation coefficients (*N* degree of decomposition) and modified details coefficients (degree of decomposition from 1 to *N*) .

The Figure 6 shows the program's interface; after image loading it is possible to choose:

- the wavelet's family used for decomposition (optional seven wavelets families with different characteristics and additional possibility of creationg own wavelet family.),
- the degree of decomposition (from 1 to 5, in this publication all images were analyzed using 2nd decomposition's degree and 1st approximation's degree),
- the discrimation threshold (setting automatically or manually),
- the type of thresholding (hard or soft, here the soft thresholding was used because it allows avoiding discontinuities in the image),
- the kind of noise to reduce (the white noise reduction was chosen).

Decomposition is applied with using Mallat's algorithm. In a method showed in this paper image is divided into four subimages: approximation and the three individual representations of the edges: vertical, horizontal and diagonal. During the noise reduction and changing the parameters values, the user has a possibility of the image preview and in a consequence visual effects evaluation. When the denoising effect is satisfactory, image is reconstructed by using the Inverse Discrete Wavelet Transform (IDWT) and it can be compared with the input image. In this paper analysis of the different types of wavelets families was made, and it was noticed that, in the case of Magnetic Resonance Imaging, the best effects are achieved using the biorthogonal wavelets.

Objective (numerical) parameters of the algorithms evaluation

The Signal to Noise Ratio and the Contrast to Noise Ratio

The Signal to Noise Ratio (SNR) is used to express how much a signal is noised [4]. It is defined as the ratio of the signal power to the noise corrupting the signal power (or as the ratio of the mean signal value to the noise standard deviation like in this

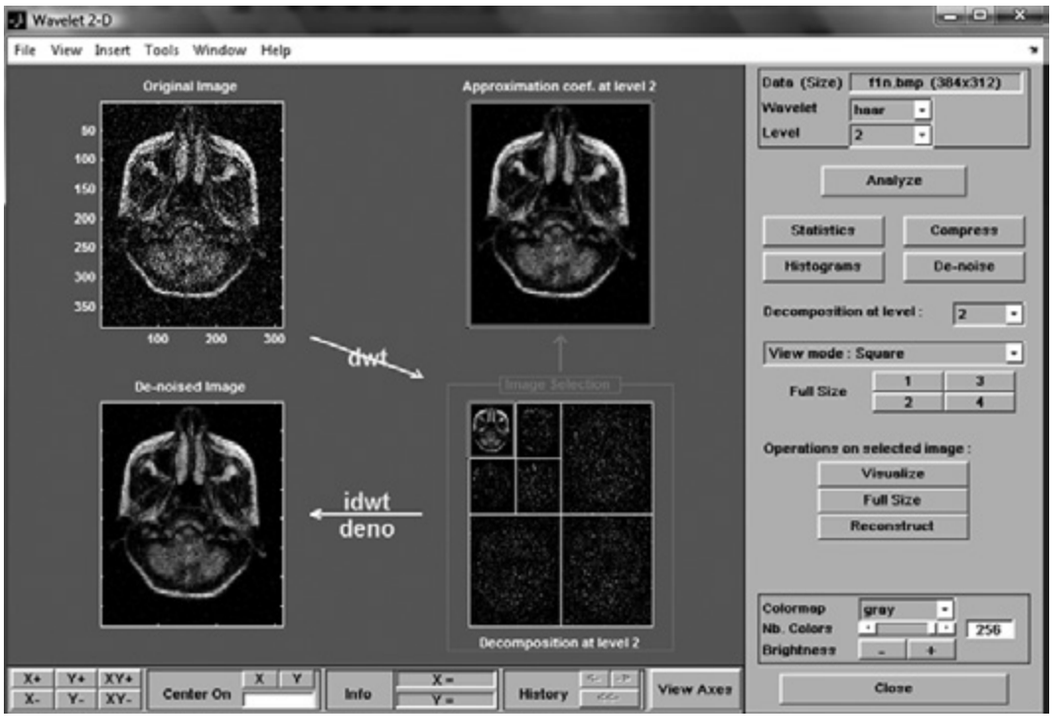


Fig. 6. Matlab Wavelet Toolbox Interface; DWT – Discrete Wavelet Transform (image decomposition), IDWT – Inverse Discrete Wavelet Transform (image reconstruction), DENO – denoising process.

paper). A ratio higher than 1:1 indicates more signal than noise. The higher the ratio, the less disturbing the background noise is.

$$SNR = \frac{\mu}{\sigma}$$

where:

μ – the mean signal

σ – the noise standard deviation (the background standard deviation)

This definition is useful for always positive variables.

$$C = S_A - S_B$$

$$CNR = \frac{C}{\sigma} = \frac{S_A - S_B}{\sigma} = SNR_A - SNR_B$$

Contrast to Noise Ratio (CNR) is a value used for defining an image quality [4]. Although this parameter is often treated as similar to Signal to Noise Ratio, the image intensity can be high even though some parts of the image are noised or blurred. That is why the image may have a high SNR value and a low CNR value in the same time. A contrast is a signal diffrence between two structures (for example two tissues). The equations defining the contrast and the CNR are:
C means contrast, S_A and S_B are signal intensities for structures *A* and *B* in the ROI (region of interest). S_A is greater than S_B so that contrast is always positive and *σ* is a standard deviation of the background.

The Table 1 shows the SNR and the CNR values in the subsequent noise reducing steps of the AWS algorithm, for the noiseless (input) and the noised image. It also presents the background standard deviation (SD) and mean signal in the processed image. Input image has of course the highest SNR, and noised image– the lowest. In each noise reducing iteration SNR increases (mainly because of the mean and SD changes). In the 6th iteration SNR is even comparable with the value for noiseless image, but the visual analysis ilustrates strong image smoothing (Figure 5), that is why SNR cannot be the only image evaluation factor. In each following iteration the image is brighter (mean value increases) and more smoothed (standard deviation decreases) despite of the input normalization. This conclusion confirms the visual images evaluation. CNR is improved (from 1st iteration), but even in the 6th iteration CNR is still lower than in noiseless image.

The method used for the image quality improvement after AWS filtration is the image merging. Merging is processed according to the equation (1). The final image is a linear combination of three images: input image α, image processed with low (β) and high (γ) *λ* value, divided by the sums of their weights coefficients. α image gets full (but noised) information about the edges and textures. Its presence in resulted image allows receiving good visual effect. In β image most of the edges and texture information are well kept, although it is still noised. In γ image only very sharp edges are kept, it approximate the general brightness distribution of the original image *λ*. Merging, by selection of best weight coefficients, allows for the optimization of the power smoothing for keeping textures and the edges in the output image. Unfortunately the best weight coefficients selection requires several tests and it depends on individual image characteristic.

	ITER	SNR	CNR
MEDICAL IMAGE	1	5,82	1,487
	2	6,67	3,022
	3	7,83	3,894
	4	8,44	5,546
	5	9,89	5,971
	6	10,71	6,115
NOISELESS IMAGE	x	10,563	7,372
NOISED IMAGE	x	5,218	1,207

Tab. 1. The SNR and the CNR changing for the AWS noise reduction, also for the noiseless and noised image (*λ*=4, $h_{(k)=}a*h_{(k-1)}$, where k means number of the iteration, a =sqrt(1.25), each h was rounded to an integer).

Although the choice of different weights values has not significant influence on the image quality the best result (the biggest SNR improvement) was received for g_α=1, g_β=8, g_γ=1. In each case the mean signal is nearly constant, that is why the best output image was received for the lowest background standard deviation. The output image σ was presented in the Figure 7.

The same SNR and CNR analysis was applied in case of the Discrete Wavelet Transform to choose the best wavelet family for the MRI applications (Table 2). In case of the MRI images the highest SNR value was received after using biorthogonal 4.4 wavelet type. In each case SNR was improved. In every case CNR was improved but also it is always less than 2. The results are similar for all wavelets families, but the best CNR was received for DB4 family, the worst – for HAAR wavelet (the highest background standard deviation).

	WAVELET FAMILY	SNR	CNR
MEDICAL IMAGE	1 HAAR	5,291	1,496
	2 DB4	6,140	1,934
	3 SYMLETS	6,069	1,913
	4 COIFLETS	6,343	1,672
	5 BIOR	6,446	1,655
	6 RBIOR	6,282	1,901
	7 DMEY	6,043	1,767
NOISELESS IMAGE	x	10,563	7,372
NOISED IMAGE	x	5,218	1,207

Tab. 2. The SNR and the CNR for the most popular wavelet families applied for the MRI medical image.

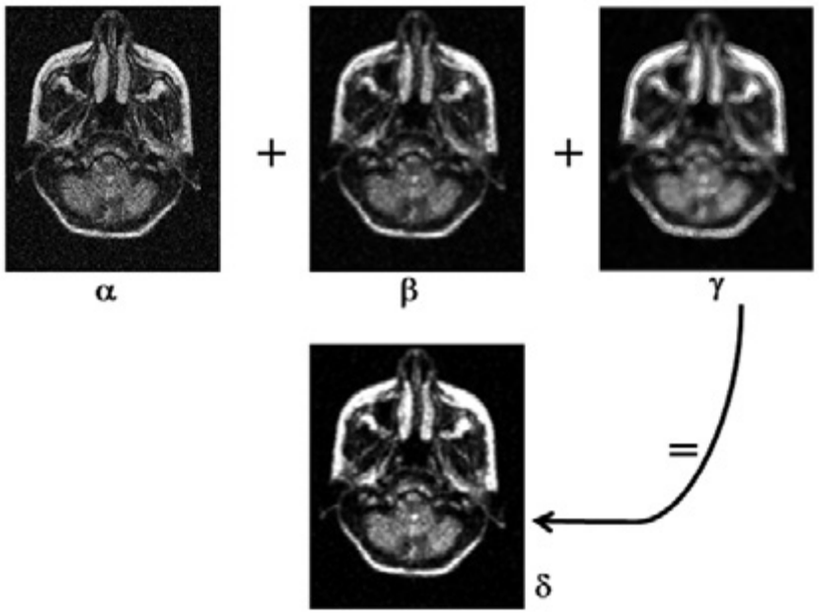


Fig. 7. The way of merging the final image (α – the input (noised) image, β – the image processed with λ=2, γ – the image processed with λ=7, weights g_α=1, g_β=8, g_γ=1, δ – the output image).

The Normal Mean value

The Normal Mean (NM) is determined for homogenous image's area without any edges. This is a very useful parameter for noise reduction evaluation [3]. It defines a relation between the variance value of image background after and before filtration.

$$NM = \frac{\mu_{filtr}}{\mu_{original}}$$

where:

μ_{filtr} – a background variance in the image after the filtration

$\mu_{original}$ – a background variance in the original image (before the filtration)

The NM values for both noise reduction methods was calculated for two kinds of the images: medical image and fantom image and given in Table 3. For the AWS algorithm NM was presented in the subsequent steps. The NM value can be viewed as a measure of the both algorithm's strength. The lower NM value, the greater image effect is achieved (greater smoothing, therefore a greater noise reduction). Bigger number of the iterations means greater image modification. While comparing the results of the AWS (3rd iteration) and the DWT algorithm the resulted values are lower for the AWS algorithm, what corresponds with the earlier conclusions about strong image smoothing by the AWS algorithm.

`	NM			
	AWS			DWT
	ITERATION 1	ITERATION 3	ITERATION 5	
FANTOM	0,877	0,574	0,399	0,687
MEDICAL IMAGE	0,815	0,501	0,379	0,718

Tab. 3. Normal mean value for the AWS (*λ*=4, $h_{(k)=}a*h_{(k-1),}$ where k means number of the iteration, a =sqrt(1.25),each h was rounded to an integer) algorithm and DWT (wavelet bior 4.4).

The mean and the standard deviation

The data presented in Table 4 prove the change of the mean value and the standard deviation (SD) value in the medical image filtered by the AWS algorithm (different I values and , different numbers of iterations). It may be noted that while the I value and the number of iterations increase, the mean value (image brightness) increases as well, and the standard deviation value (image smoothing and homogeneity) decreases in the same time. Although numerically changes are not very high, visual effect is significant like it was already shown in the Figure 5.

The changes of the mean value and of the standard deviation value in the image filtered by the DWT algorithm for two different medical images: medical one (without noise, noised and after DWT filtration) and for the medical fantom image (noised and after DWT filtration, noiseless image was not available) can be read from the data presented in Table 5. It is shown that just like in case of the AWS algorithm the filtration slightly increases the mean and the standard deviation value.

	λ COEFFICIENT FOR THE MEDICAL IMAGE								
	λ=1			λ=4			λ=7		
	ITERATION			ITERATION			ITERATION		
	1	4	6	1	4	6	1	4	6
MEAN	63.33	64,11	67,25	64,08	67,16	78,23	64.28	67.35	83.54
STANDARD DEVIATION	9,06	9,22	9,86	8,98	9,39	9,37	9.03	9.89	10.74

Tab. 4. The mean and the standard deviation change in the images denoised by AWS algorithm ($h_{(k)}=a*h_{(k-1),}$ where *k* means number of the iteration, *a =sqrt(1.25),*each *h* was rounded to an integer).

	IMAGE				
	MEDICAL IMAGE			FANTOM IMAGE	
	WITHOUT NOISE	NOISE	DWT	NOISE	DWT
MEAN	51,42	62,09	70,35	82,97	88,03
STANDARD DEVIATION	8,97	9,59	9,25	22,38	19,08

Tab. 5. Mean and standard deviation evolution in the images denoised by DWT (wavelet bior 4.4).

Histograms

A change in the intensities of the pixels caused by the action of both algorithms may be presented using histograms. The Figure 8 shows changes in the image histogram after its noising and for the both methods of noise reduction. In the original noiseless image there is a significant background peak (low intensity). The noising process removes that peak (smoothing the image), but its recovery is not necessary. For the diagnostic value, light pixels are much more important (they determine the image contrast). Moreover, light pixels are more noisy (even for the constant noise, its percentage contribution in the signal is larger for light pixels). Histograms for both methods confirm earlier conclusions about the images brightening (shifting the main peak in the direction of the higher pixel intensities) and smoothing (less half width of the peak) after their filtration.

Detection of edges

Detection of edges detection is an extremely important part of the image quality improvement In this procedure some characteristics important for the image diagnostic utility should be taken into consideration. The most important is the lack of the false edges, detected because of the presence of noise, which must be reduced. In the Figures 9 and 10 the differences in edges detection and the false edges reducing are shown for both methods. The Figure 9 presents the edge detection in case of the noised fantom image. The simple edge detection filter was used. After its application all the edges were sharpened. For the AWS algorithm the edge detection in different iterations were illustrated. It is shown that in case of strong smoothing, in the subsequent iterations edges are getting blurred and do not correspond with the edges in the input image. Edges detected in the 1st and 3rd AWS iteration are similar but not identical. In the 3rd iteration not existing edges were detect (resulting image shows double edges). The slight difference between edges detection in the input image and the image processed ny AWS algorithm in the early iterations, as well as the fact that all the real edges were correctly identified are the consequences of the small number of the strong edges with very high intensity gradient (high difference in the pixels intensities on the both edges sides). In case of using the DWT method few wavelets families were tested. The best results were once again observed for biorthogonal 4.4 type wavelet. The results are slightly better than the for the simpler wavelets families like db 2 and coifflets 4 (apart from the real, anatomical edges the false artificial edges are detected as well).

Analysis of the edges detection in the medical image(with bigger number of the edges than the fantom image), shown in the Figure 10 is much more interesting case. After its application for the noiseless image, all the anatomical edges were sharpened. Using the same method for noised image shows a big number of the false, redundant edges. The image is unreadable and diagnostically useless. That is why both filtration methods were applied. After the AWS filtration all the redundant edges were removed. Unfortunately the AWS algorithm reduced also real, anatomical edges; undetecting all the real edges is a significant disadvantage of this method. Changing the value of the I value may influence on the degree of anatomical edge detection, but it requires several attempts to optimize. On the other hand, the DWT method detected all the needed edges, but also some of the false. The degree of redundant edges detection depends on wavelets family. It is necessary to take into consideration the best wavelets family for specific cases.

The basic advantage of using the AWS algorithm is good preserving of the edges continuity after smoothing. This method also removes the false edges created by the image noise. By using the AWS algorithm it is possible to receive single pixel edges without changing their position or their course. It should be also noted that it has only two input parameters. Unfortunately this method has also some disadvantages. For the high I value (necessary for strong smoothing) there is a possibility of losing information about small structures (strong decreasing the I value causes false contours).

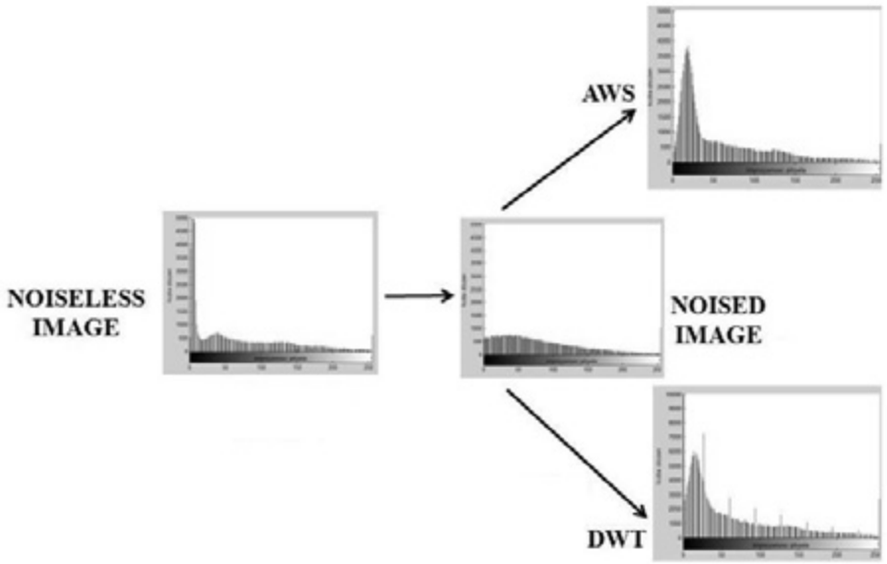


Fig. 8. Changing of the image histogram during the noise reduction in the medical image (λ=4, $h_{(k)}=a*h_{(k-1),}$ where *k* means number of the iteration, *a =sqrt(1.25),* each *h* was rounded to an integer, bior4.4 wavelet).

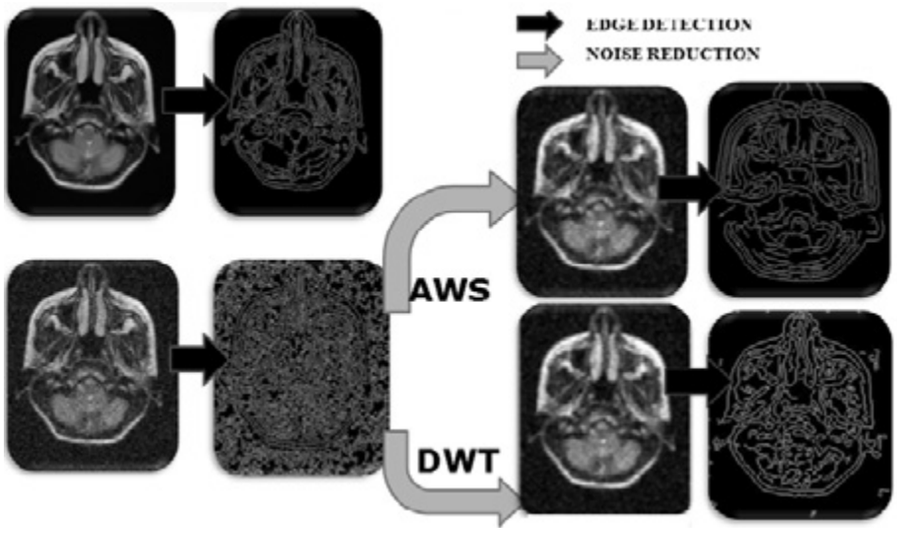


Fig. 9. Keeping the edges after noise reduction for the fantom image (λ=4, $h_{(k)}=a*h_{(k-1),}$ where *k* means number of the iteration, *a =sqrt(1.25),* each *h* was rounded to an integer).

Summary and conclusions

The most serious disadvantage of the AWS method is the time consumption. The algorithm is iterative and the weights are calculated for each pixel within the surrounding, which is why the computational complexity of the algorithm is very high. It takes also much time because of the nessesity to perform the tests guaranteeing the best results.

A huge advantage of the DWT algorithm is the possibility of filtering in *k*–space (not possible for the AWS method). Comparing these two methods it is necessary to mention linear computational complexity for the DWT, what gives shortening calculation time. Both of them keep the edges satisfactorily. In the case of the DWT there might be some optimization problems (signals require an individual approach). The DWT algorithm does not remove all the false edges, what reduces the image quality. What is more this method requires bigger number of the input parameters than the AWS algorithm.

Adaptive filter methods like AWS algorithm give good visual effects (by using the merging images procedure). For high noise (low SNR value) small artifacts can be highlighted as the points with the equal value (correction by increasing λ parameter), and the best results are guaranteed for λ belonging to range from 2 to 4. Detailed analysis of the influence of λ value on the results in Magnetic Resonance Imaging shows that the best results are available for λ=4. The AWS algorithm gives best results

for images with primary noise reduction (then big number of the iterations is not required). The AWS algorithm is keeping the edges, but the average image brightness slightly increases (despite of the image normalization). Disadvantageous process of too strong smoothing the image and blurring the edges make the algorithm not very effective for too big numbers of iterations. The algorithm can be applied to improve MRI images in the case of medium noise and low Signal to Noise Ratio.

The Discrete Wavelet Transform (DWT) increases image brightness and makes it more homogenous. It keeps all the anatomical edges, unfortunately also some of redundant edges remain. Its time consumption is significantly lower than for the AWS.

The presented results prove that the further application of the AWS and the DWT methods, to improve MRI images are useful but their application needs further studies.

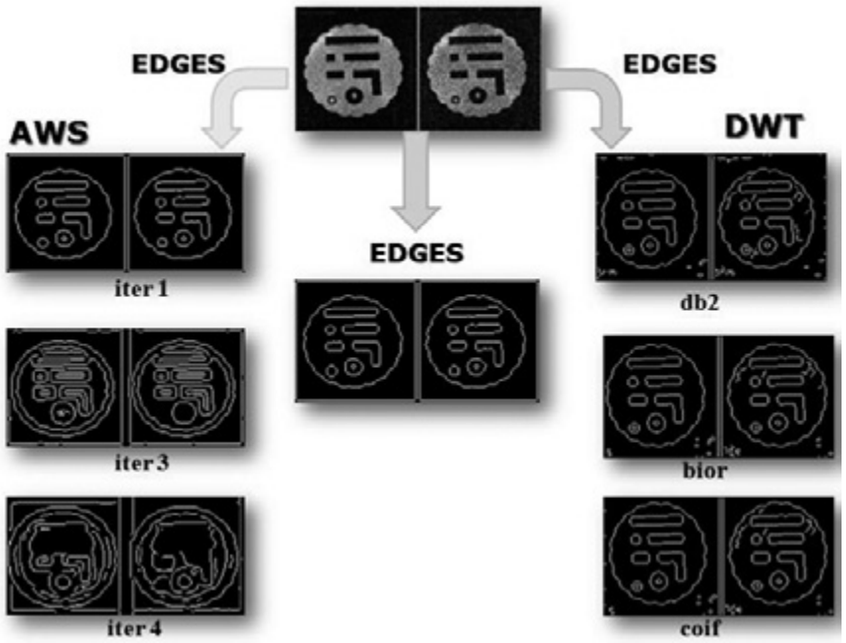


Fig. 10. Keeping the edges after noise reduction for the medical diagnostic image (λ=4, $h_{(k)}=a*h_{(k-1)}$, where *k* means number of the iteration, *a =sqrt(1.25),* each *h* was rounded to an integer, iter 3rd, biorthogonal 4.4 wavelet type).

Acknowledgment

Support from the Ministry of Science and Higher Education through the statutory funds for the Faculty of Physics and Applied Computer Sciences, AGH University of Science and Technology Cracow, is acknowledged.

Image references

1. The brain pictures T_2-weighted at 1,5 T are the Author (J.Ś.W.) property.
2. The Authors express their grattitude to MSc. Bartosz Proniewski for the fantom pictures at 0.088 T.

References

1. Polzehl J., Spokoiny V.G.: Adaptive Weights Smoothing with applications to image restoration. *J. R. Stat. Soc. Ser. B Stat. Methodology* 2000, 62: 335-354.
2. Polzehl J., Spokoiny V.G., Propagation–Separation Approach for Local Likelihood Estimation. *Probability Theory and Related Fields* 2006, 135: 335- 362.
3. Kolecki J., Badurska M., Despeckling of synthetic aperture radar images using propagation–separation approach for local likelihood estimation. *Archives of Photogrammetry, Cartography and Remote Sensing* 2008, 18a: 251-261.
4. Alexander M.E., Baumgarter R., Summers A.R.: A Wavelet based method for improving signal–to–noise ratio and contrast in MR images. *Magnetic Resonance Imaging* 2000, 18: 169-180.
5. Kolecki J., Wróbel A.: Edge detection using adaptive smoothing of digital images. *Archives of Photogrammetry, Cartography and Remote Sensing* 2008, 18a: 263-272.
6. Białasiewicz T.: Falki i aproksymacje, Wydawnictwo Naukowo-Techniczne: Warszawa 2000, pp. 204-218.

Figure captions

Fig. 1. Pixel surrounding definition.

Table captions

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 67-71

ANALYSIS OF STRESS DISTRIBUTION IN HIP JOINT AFTER IMPLANTATION OF WELLER STEMS

Wojciech Więckowski

Zakład Inżynierii Biomedycznej, Politechnika Częstochowska, Częstochowa, Dąbrowskiego 69, Poland, wieckowski@iop.pcz.pl

Summary: This study presents the results of numerical calculations of stress distribution in a model of healthy femur and after implantation of Weller II cement prosthesis. The obtained stress distribution allow for drawing conclusions about the processes of bone remodelling in the area of implant and for forecasting of effectiveness of implantation. The simulations were carried out by means of FEM for three types of stem material: CoCrMo alloy and titanium alloys of Ti6Al4V and Ti35Nb5Ta7Zr. The calculations revealed that implantation of prosthesis stem leads to unloading of prosthesis bone tissue, depriving it of natural stress which occurs in healthy femur, which might lead to loosening of the implant.
Keywords: Weller endoprosthesis, hip joint alloplasty, computer simulation

Introduction

In the case of advanced degenerative changes in hip joint, cement or noncement (either partial or total) hip joint replacement seems to be the only method which allows for restoration of the functions in damaged joint.

An essential determining factor of the success in hip joint replacement is the architecture of prosthesis. A number of studies have demonstrated that suitable adaptation of stiffness of prosthesis to the adjacent bone, obtained through selection of a suitable shape of stem (its curvature, dimensions, mechanical properties), considerably impacts on the method of transfer of load from implant to the bone. Redistribution of stress in prosthesis tissue, discontinuities and concentration of stress lead to occurrence of changes in bone tissue in femur (remodelling) and, in consequence, to longterm loosening of the prosthesis [1, 5, 7, 9, 12].

Design solutions for new types of prostheses made of modern biomaterials have brought significant improvement in the quality of hip joint replacement. However, problems connected with the stability of implants remain unsolved. This is confirmed by the results of clinical studies which have reported that despite substantial progress in the domain of hip joint design, a considerable percentage of implantations end in failure [2, 10, 11, 14].

Determination of an impact of design of prosthesis on the way the load is transferred and on stress distribution in bone-implant system, and, in consequence, on its life, is possible through application of modern computer systems. The results of numerical simulations allow for explanation of the reasons for destruction processes which occur in the area of implanted joint and for forecasting of effectiveness of implantation [3, 14].

Aim and scope of the study

Weller cement prosthesis of type II was one of the most frequently implanted hip joint prostheses in Poland in the last years of the recent century. Stems in these prostheses are curved and feature a flange which prevented from prosthesis deepening into the bone. The unsatisfactory results of Weller cement hip joint replacement presented in the literature confirm design imperfectness in this type of prosthesis stem [2, 10, 13].

The goal of the numerical investigations was to determine the effect of material parameters and design features of femur component in hip joint prosthesis on the processes connected with its biofunctionality. Numerical calculations for stress distribution were carried out in threedimensional model of healthy femur and after Weller II cement prosthesis stem. Comparative analysis of the results of calculations for characteristic crosssections of bone tissue in the model of healthy bone and bone-cement-implant system (Fig. 1c) was also made.

Method

Realization of the tasks required development of a model of femur with the areas taken by cortical bone and spongy bone and the model of the system of bone-cement-implant. Geometrical features of femur were based on the model of Standardized Femur developed by Instituti Ortopedici Rizzoli in Bologna (Fig. 1a). Geometrical model of prosthesis stem was developed based on the data obtained from measurement of Weller II stem (NA 15 K) by means of coordinate measuring machine (Fig. 1b). Development of the model of bone-cement-implant required determination of position of the prosthesis in relation to femur and modelling of cement layers on the surface of prosthesis which allow for integration of implant with the adjacent bone (Fig. 1c). In order to build discrete models, 20 node-type 3D Solid elements were used. Total mechanical connection of prosthesis with the layer of bone cement and cement with the area taken by the bone were assumed. The phenomenon of contact on the resistant surface of the flange which occurred as a result of removal of femur head on which 3D Contact elements were generated.

Material model of bone-cement-implant system, in which linear elastic properties were assumed for each of the elements, was developed based on the results from empirical studies presented in the literature on mechanical properties of bone and research works focused on hip joint implant modelling [4, 8]. Numerical models took into account three types of prosthesis stem material: CoCrMo, Ti6Al4V and Ti-35Nb-5Ta-7Zr alloys. Strength parameters adopted for calculations are presented in Table 1.

A simplified spatial model of load, corresponding to one-leg standing. The model took into consideration the effect of external forces which were applied on the prosthesis head: P1=2.47BW (BW – body weight) and greater trochanter in femur P2=1.55BW [3]. ADINA software was used for calculations.

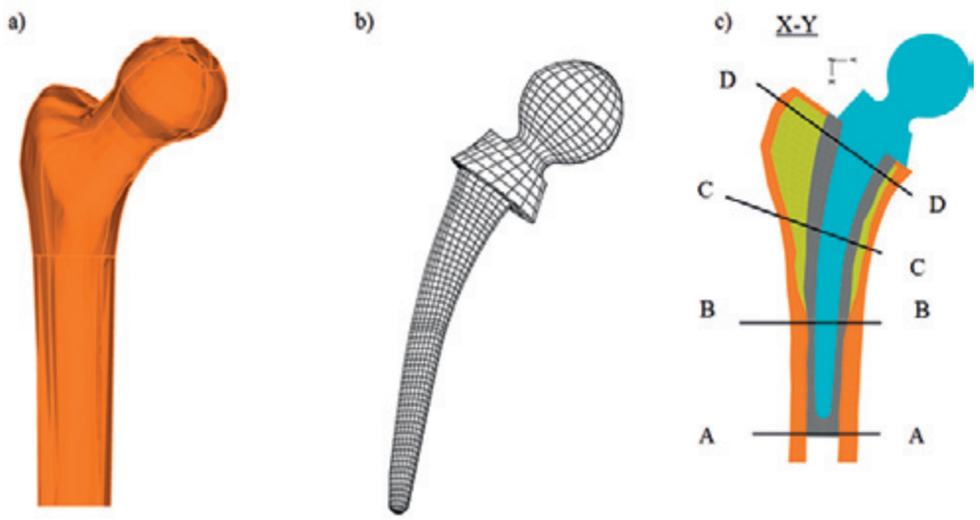


Fig. 1. Geometrical model of a) healthy femur, b) Weller II stem, c) bone-cement-implant system.

	cortical bone	spongy bone	cement	CoCrMo	Ti6Al4V	Ti-35Nb-5Ta-7Zr
Young's modus [MPa]	1.70×10^4	0.10×10^4	0.24×10^4	2.00×10^5	1.10×10^5	0.55×10^5
Poisson's ratio *v*	0.35	0.40	0.33	0.30	0.30	0.30

Tab. 1. Mechanical properties of bone tissue and prosthesis stem material [4,6,8].

Results

Distribution of reduced stress σ_{zr} in cortical and spongy tissues in the model of femur before and after implantation of Weller II prosthesis stem are presented in Fig. 2a and Fig. 2b, respectively.

In healthy femur, load is transferred through femur head to cortical tissue of proximal metaphysis, from where their peripheral transfer to further part of the femur body occurs.

A general view presents the location of maximal values of stress. The presented results of calculations reveal that it occurs in medial part of femur at the level of lesser trochanter and amounts to $\sigma_{zr\ max} = 46$ [MPa]. The highest values of reduced stress of $\sigma_{zr} = 25 \div 46$ [MPa] can be observed throughout the length of body of the femur, in internal and lateral areas.

Implantation of Weller stems leads to the reduction in the value of reduced stress in the whole area of cortical tissue com-

pared to those which occur in the model of healthy femur. The highest decrease in the value of reduced stress (28%) was observed in proximal part of femur. In body of the femur below the location of implant's end, constant distribution of reduced stress can be observed, with reduced stress reaching 90% of the values which occur in femur before implantation in this area.

Change in stem material, from CoCrMo into titanium alloy causes rise in reduced stress in the area of bone tissue. The effect of stem material on the value of reduced stress can be observed in particular on the internal side of proximal metaphysis in the femur. In further part of body of the femur, no significant effect of type of stem material on changes in the values of reduced stress was observed.

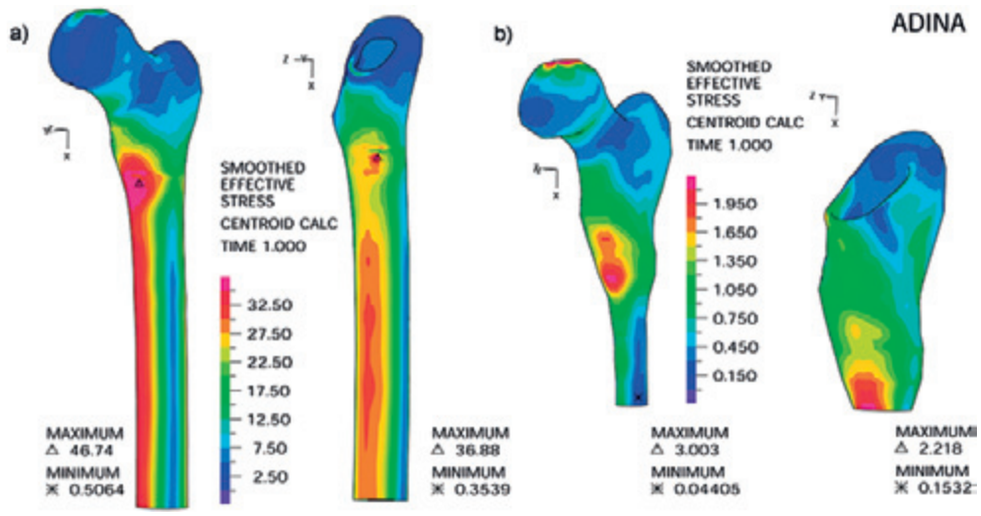


Fig. 2. Distribution of reduced stress σ_{zr} [MPa] in the area of cortical (a) and trabecular (b) bone in the model of healthy femur and after implantation of Weller stem.

Comparison of maximal values of reduced stress in the analysed cross-sections in the area of cortical and spongy tissue of the bone model after implantation of Weller stems is presented in Fig. 3.

Distribution of reduced stress σ_{zr} in Weller prosthesis stem and in the layer of cement, whose life is of significant effect on stable fixation of the stem is presented in Fig. 4.

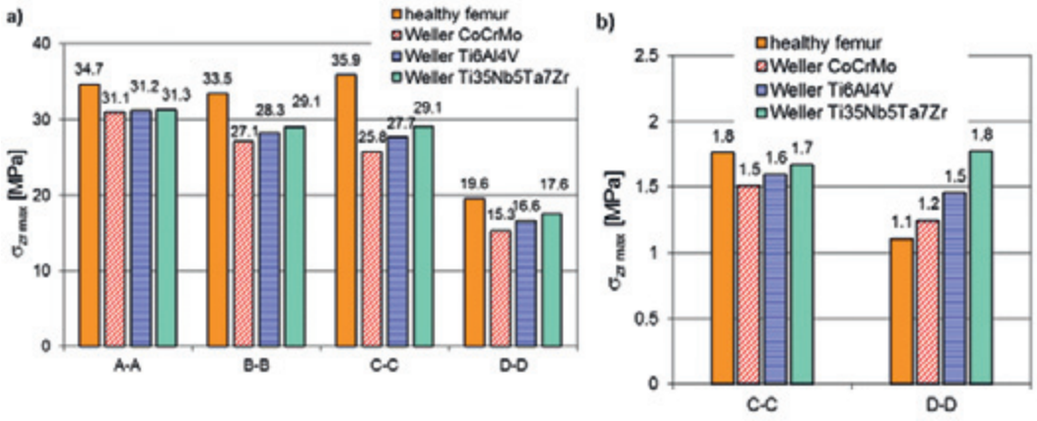


Fig. 3. Maximal values of reduced stress σ_{zr} [MPa] in cross-sections of the area of: a) cortical tissue, b) spongy tissue.

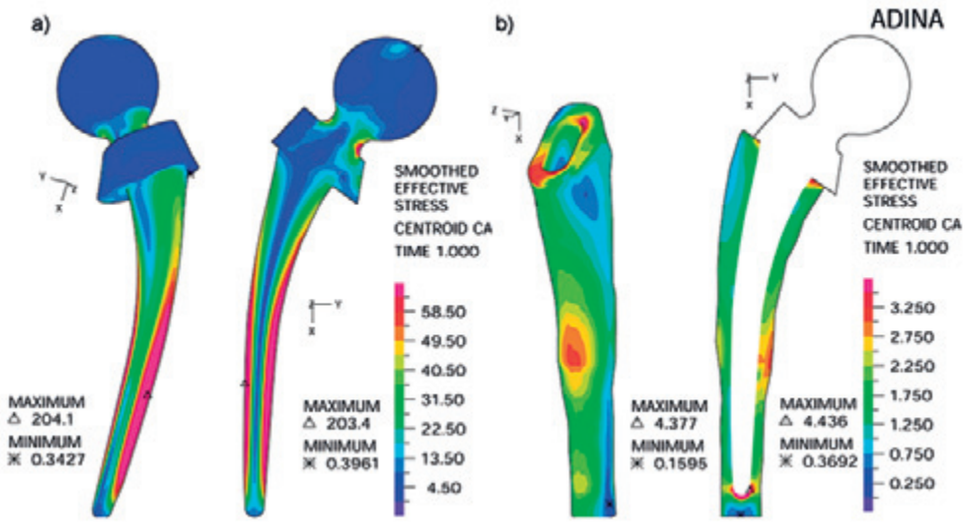


Fig. 4. Distribution of reduced stress σ_{zr} [MPa] in prosthesis stem (a) and cement layer (b) in the model of bone-cement-prosthesis Weller II (stem material: CoCrMo).

The highest values of reduced stress at the level of $\sigma_{zr} = 2.5 \div 4.4$ [MPa] are observed in the layer of cement located on the medial part of prosthesis stem and result from the mode of load transfer from implant to the bone around the implant. Reduction in stiffness of the stem is accompanied by rise in stress in cement coating located at internal side of proximal part of implant blade, which might lead to cracking of cement layer and, in consequence, to loosening and damage to the prosthesis.

Summary

The presented results of calculations confirm occurrence of significant changes in load transfer in femur caused by implantation of prosthesis stem. Prosthesis implantation leads to unloading of bone tissue in the area of implant, depriving it of proper stress distribution typical of natural healthy hip joint, which stimulates bone mass.

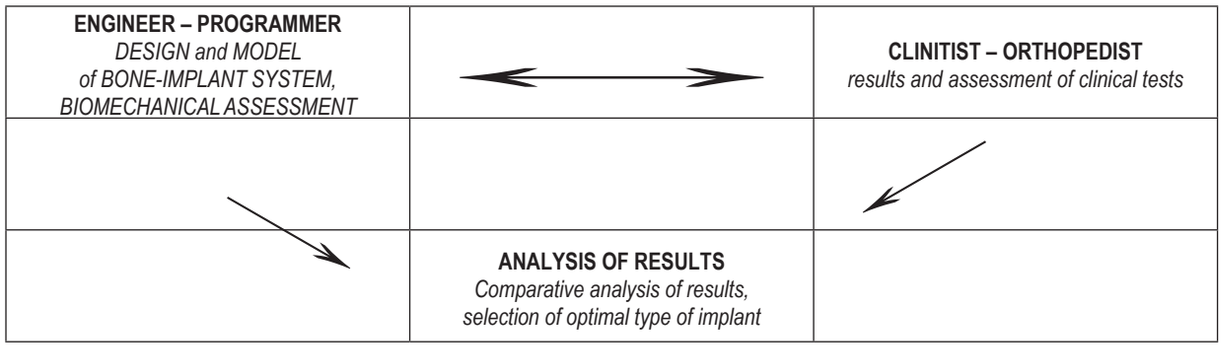
The numerical calculations presented in this study revealed that in the case of application of flange-type Weller II stems, considerable part of load is transferred to the bone through resistant surface of the flange, which might lead to excessive and undesirable load on flange-bone contact surface. As results from clinical studies, a phenomenon of osteolysis under the prosthesis flange can be observed, especially if the stem is deepened into the femur during use, which leads to large unit pressures on the area of bone tissue under the flange. If the prosthesis is properly designed and well-fitted in the femur's hole, stress under the flange is insignificant and the osteolysis phenomenon occurs only to a small extent. Load transferred through the flange leads to occurrence of compressive stress in the bone, which prevents from bone structure remodelling.

In the group of the investigated types of materials, the most favourable stress distribution in the bone was observed for stems made of titanium alloys, which confirms their usefulness for hip joint replacement. Furthermore, good corrosion resistance and lower density compared to CoCrMo justify application of these materials for hip joint prosthesis stems. In cement-based hip joint prostheses, reduction in Young's modulus for stem material leads to the rise in stress in proximal area of cement layer, therefore, in the case of cement fixation of implant, use of 'flexible' stems is purposeless.

References

1. Balin A.: Mechaniczne i materiałowe uwarunkowania stabilności trwałości endoprotezy stawu biodrowego. *Inżynieria Materiałowa* 1997, No. 2: 44-52.
2. Biliński P. J., Wolański R., Gumański R., Mątewski D.: Wyniki leczenia zmian zwyrodnieniowych stawu biodrowego endoprotezoplastyką cementową Wellera. *Chirurgia Narządów Ruchu i Ortopedia Polska* 1996, LXI, Supl. 3A: 17-20.
3. Dietrich M., Kędzior K.: Design and manufacturing of the human bone endoprostheses using computer-aided systems. *Journal of Theoretical and Applied Mechanics* 1999, No. 3, Vol. 37: 481-503.
4. Krzesiński G., Zagrajek T.: Modelowanie własności mechanicznych kości. *Biology of Sport* 1998, Vol. 15, Suppl. 8: 238-243.
5. Kusz D.: Porównawcza ocena zmian mineralnych w kości udowej po endoprotezoplastyce cementowej stawu biodrowego typu Weller i Centrament. *Chirurgia Narządów Ruchu i Ortopedia Polska* 1999, 64(4): 407-414.
6. Marciniak J.: Biomateriały. Wydawnictwo Politechniki Śląskiej: Gliwice 2002.
7. Okrajni J., Kusz D., Żmudzki J., Kopka W.: Mechaniczne czynniki determinujące stan tkanki kostnej w otoczeniu implantów. *Acta of Bioengineering and Biomechanics* 1999, Vol. 1, Suppl. 1: 345-348.
8. Reilly D.T., Burstein A.H.: The Mechanical properties of cortical bone. *The Journal of Bone and Joint Surgery* 1974, Vol. 56-A, No. 5: 1001-1022.

9. Sumner D.R., Turner T.M., Igloria R., Urban R.M., Galante J.O.: Functional adaptation and ingrowth of bone vary as a function of hip implant stiffness. *Journal of Biomechanics* 1998, 31: 909-917.
10. Ziółkowski W., Kamiński M., Doerr-Winogradow K., Ostojski R.: Radiologiczne i histologiczne zmiany w jałowym obluzowaniu udowej części endoprotezy stawu biodrowego według Wellera. *Chirurgia Narządów Ruchu i Ortopedia Polska* 1996, LXI, Supl. 3A:121-125.
11. Koczy B., Stołtny T., Pająk J. i in.: Realloplastyka komponentu panewkowego endoprotezy stawu biodrowego bez użycia mechanicznych systemów wzmacniających. *Chirurgia Narządów Ruchu i Ortopedia Polska* 2009, 74(4): 207-213.
12. Ryniewicz A., Madej T.: Analiza biomechaniczna w stawie biodrowym po całkowitej alloplastyce z różnymi wariantami endoprotez cementowych. *Mechanika w medycynie* 2010, 10: 253-264.
13. Cwanek J., Liubimov V.: Przydatność parametrów geometrii powierzchni do oceny stopnia zużycia trzpienie endoprotez Wellera, *Mechanika w medycynie* 2010, 10: 45-51.
14. Będziński R., Ścigała K.: Biomechanika stawu biodrowego i kolanowego, Biocybernetyka i inżynieria rehabilitacyjna, t. 5. Akademicka Oficyna Wydawnicza EXIT: Warszawa 2004.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 73-78

NOVEL DETECTOR SYSTEMS FOR THE POSITRON EMISSION TOMOGRAPHY

P. Moskal, P. Salabura, M. Silarski, J. Smyrski, J. Zdebik, M. Zieliński

Institute of Physics, Astronomy and Applied Computer Science,
Jagiellonian University, 30-059 Cracow, Poland

Abstract: In this contribution we describe a novel solution for the construction of Positron Emission Tomograph. We present the device allowing for determination of the impact position as well as time and depth of interaction of the annihilation gamma quanta. The device is comprised of scintillation chamber consisting of organic scintillators surrounding the body of the patient. We discuss two possible solutions: (i) the tomograph built out of scintillator strips, and (ii) the tomograph built out of the scintillator plates. The application of the fast scintillators will enable to take advantage of the difference between time of the registration of the annihilation quanta. The invented method will permit to use a thick layers of detector material with the possibility of measuring the depth of the gamma quantum interaction (DOI) and the determination of their time of flight (TOF), and will allow for increasing the size of the diagnostic chamber without a significant increase of costs. The method is a subject of two patent applications [1, 2] which are based on the techniques used in the particle physics experiments [3, 4].
Keywords:

Introduction

Positron Emission Tomography (PET) is at present the most technologically advanced diagnostic method that allows for non-invasive imaging of physiological processes occurring in the body. It plays a fundamental and unique role both in medical diagnostics, as well as in monitoring of effects of therapy in particular in oncology, cardiology, neurology, psychiatry or gastrology. PET tomography constitutes also an effective tool to investigate the functioning of the brain. PET permits to determine the spatial and temporal distribution of concentrations of selected substances in the body. To this end, the patient is administered pharmaceuticals marked with radioactive isotope emitting positrons. Since the rate of assimilation of marked pharmaceuticals depends on the type of the tissues, sections of the diseased cells can be identified with high accuracy, even if they are not yet detectable via morphological changes. The method is proved to be extremely effective in particular in locating and diagnosing of cancer metastases.

All known matter including the body of the patient is built out of electrons, protons and neutrons. The PET uses the fact that the electron and positron annihilate while contact with each other and their mass is converted to energy in the form of gamma quanta. Most frequently these are two gamma guanta flying against each other along the line with an exactly defined energy equal to 511 keV. PET permits to locate the radioactive marker by the use of radiation detectors, allowing to reconstruct the direction of flight of annihilation quanta. Radiation detectors are usually arranged in layers forming a ring around the diagnosed patient (see Fig. 1). The set of reconstructed lines (referred to as Line of Response: LOR) constitutes the basis for the reconstruction of the thomografic image which reflects the distribution of the density of the radiofarmaceutic in the body of the patient. This technique allows to investigate the physiological processes involving the radiofarmaceutics since PET permits to obtain few tens of images within a minute.

The typical radiation dose in examination with PET amounts to about 7 mSv [5]. This is comparable with similar doses obtained by the patient in other diagnostics methods. It is in the order of the avarage yearly dosis due to the natural radiation sources as cosmic rays, concentrations of radon in the air etc. It has therefore no negative influence on the patient and even could have a positive implications in the functionning of the immune system due to the well established hormesis mechanism [6, 7].

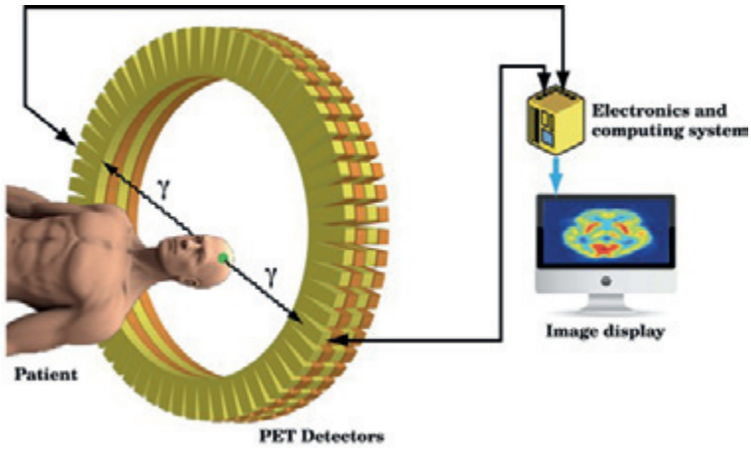


Fig. 1. Schematic illustration of the PET tomography.

A natural limitation of the sharpness of the PET image is given by the fact that positrons annihilates predominantly after its kinetic energy is decreased to the values close to zero which is typically few milimeters far from the nucleus from which it was emitted. As regards the detection technique among main factors limiting presently achievable accuracy are the problem of unknown depth at which gamma quantum reacts (referred to as DOI – depth of interaction), the problem of insufficient time resolution of non-organic detectors preventing them from the effective usage of the time difference between the arrival of the gamma quanta to the detectors (TOF – time of flight [8]) and, finally, present solutions are impractical to perform a tomographic image of the whole body at the same time.

In the following sections the presentation of the invented detector systems will be preceded by the description of the disadvantages in the tomografic image reconstruction caused by the unknown DOI and possible gains which could be achieved with the good resolution for TOF determination.

Depth of Interaction

Currently, all commercial PET devices use inorganic scintillator materials as radiation detectors (usually BGO cristals). The energy of gamma quantum hitting the scintilator can be transfered partially or entirely to an electron of the material, which then produces flash of lights through ionization and deexcitation of atoms or molecules of the scintillator. These flashes are then converted to electrical pulses by photomultipliers connected to the scintillators.

As it was shown in Fig. 2a scintillating crystals, made usually in size of about 5cm x 5cm and with thickness of 2.5 cm, are additionally blazed into smaller pieces with dimensions of 0.5 cm x 0.5 cm separated form each other with reflecting material. The end of each scintillating module is connected to photomultipliers which convert light into electrical impulses (Fig. 2b). Distribution of the amplitude of this impulses permits to determine, with the accuracy equal to the size of the small unit, the position where the

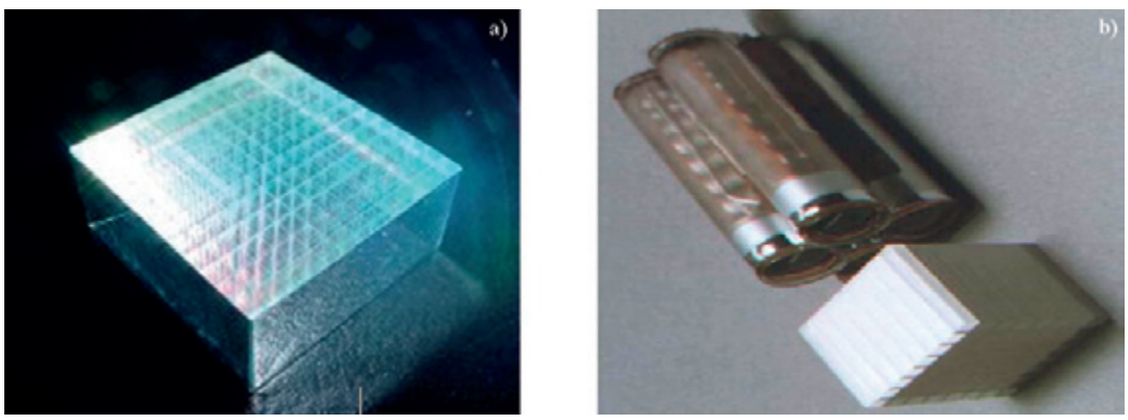


Fig. 2. a) Shape of a single scintillating module used to detect gamma quanta in PET scanners.
b) Typical configuration of photomultipliers which register light from a single scintillating module used in PET scanners.
The figures are adapted from [9, 10].

gamma quantum reacted. Therefore, in the further analysis, to determine the LOR line, one assumes that the quantum was absorbed in the middle of the unit. This assumption is one of the most important contribution limitting the resolution of PET images. Problem of unknown depth where the gamma quantum reacted is known in the literature as DOI (depth of interaction).

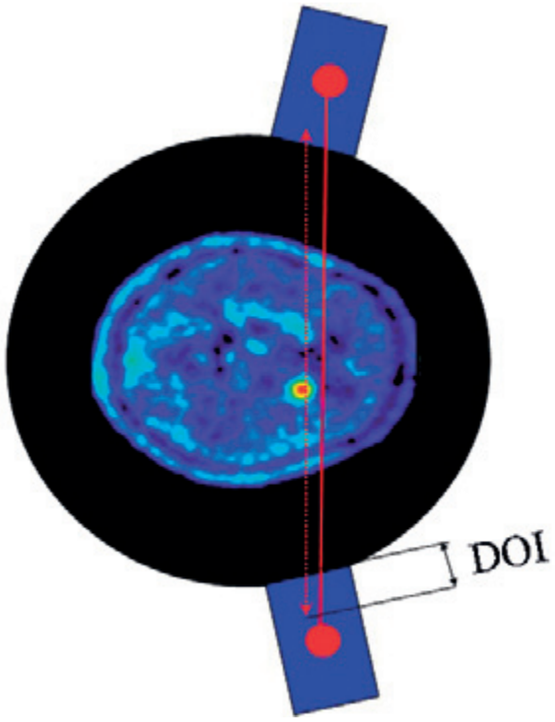


Fig. 3. Schematic illustration of the error in reconstruction of the LOR line made due to the unknown depth of interaction of the gamma quantum.

In Fig. 3 we present, in big simplification and exaggeration, the error in the determination of the LOR line made due to the unknown DOI. The dashed line shows the real path of flight of the gamma quantum, while the solid line represents the LOR reconstructed assuming that the signals emerged in the middle of the detection module. Distortions are the greater, the farther from the axis of the tomograph the annihilation occured, and the larger is the scintillator module. Therefore, the determination of the DOI could improve a lot the resolution far frorm the axis leading to better imaging of whole body. Moreover, it could allow to use a thicker scintillators improving the efficiency of the measurement. However, to our knowledge, at present none of commercially produced tomographs is able to measure the DOI.

Time-of-Flight

The other way to improve the resolution of the tomographic image is determination of the annihilation point on the LOR line based on measurements of the time difference between the arrival of the gamma quanta to the detectors. In the literature this technique is known as TOF (time of flight), and tomographs which use the time measurements are termed PET-TOF. In the idealized case, as it is shown in Fig. 4, measurement of the time difference between arrival of the quanta (t2-t1) could allow calculation of the annihilation point relative for example to the center of the LOR line denoted in Fig 4. as Δx. In practice, due to the finite resolution of the time measurement, it is possible to determine only a range along the LOR in which the annihilation occured which also improves the resolution of PET images.

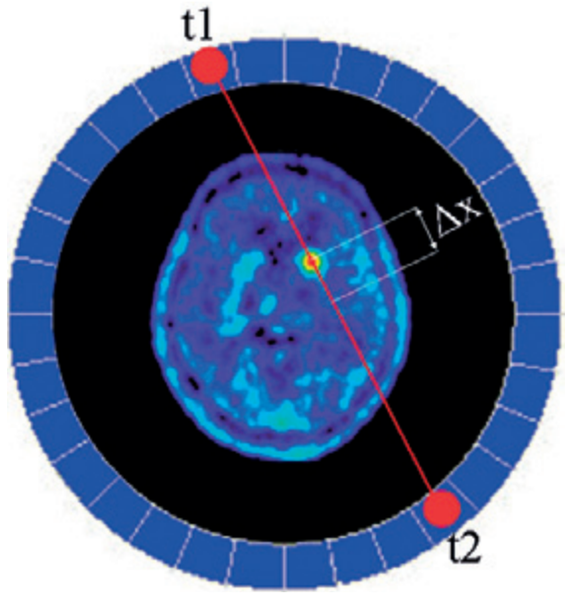


Fig. 4. The idea of PET-TOF; Δx = (t2 – t1) c/2.

Scientist have tried to use the time of flight of gamma quanta in the PET tomography since 1980 [11]. But so far nobody has obtained a significant improvement using TOF method, mainly because of the inorganic scintillators applied in PET scanners which give slow impulses. In 2008 a prototype made by SIEMENS achieved the time resolution of about 550 picoseconds which corresponds do the spatial resolution along the LOR line amounting to 8 cm [11]. About one order of magnitude better time resolutions can be achieved with organic plastic scintillators, which were so far not used due to low density and a small atomic numbers of the elements constituting the material. Fast plastic scintillators are composed mainly of carbon and hydrogen. Small atomic number corresponds to small probability that gamma quanta transfer all its energy to the electron in the scintillator through the photoelectric effect. Moreover, small density implies a small efficiency for the detection of gamma quanta. The efficiency could be improved by increasing the thickness of the scintillator, but on the other hand it would decrease the resolution of the image due to DOI problem. However, novel methods presented in the next sections allow to increase the thickness of the detector and at the same time to determine the depth of the interaction of the registered gamma quantum. In addition due to the large solid angle covered by the new PET construction the decrease of the detector efficiency will be compensated by the increase of the acceptance. A small efficiency for the photoelectric effect in organic scintillators worsens the image quality due to a low ability

to distinguish between quanta reaching the detector directly and quanta rescattered in the body of a patient. This dra wback will be compensated by (i) the selection of only these events for which the energy deposited in the scintillator corresponds to the range close to the maximum of energy which can be transfered to the electron via the Compton scattering process, and by (ii) taking advantage of the good timing of the organic scintillators allowing for the effective usage of the TOF technique.

In both below discussed methods the reconstruction of the interaction point of the gamma quantum in the scintillator material is reconstructed based predominantly on the time distribution of signals measured at various parts of the detector.

Novel Solution: Strip PET

In the "strip" PET the test chamber is formed from organic scintillator strips contracting the cylinder. Light signals from each strip are converted to electrical signals by two photomultipliers placed at opposite edges of each scintillator strip. The scheme of the single detector module is shown in Fig. 5. To determine the impact position of the gamma quantum the time difference between signals from both ends of the strip is used and the time of the interaction of the gamma quantum in the strip is determined as an arithmetic mean of the times measured on both edges of the scintillator.

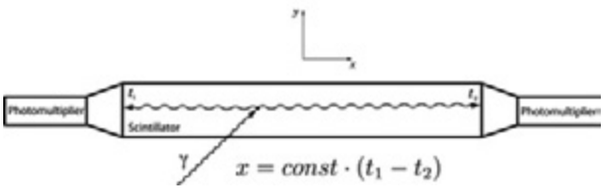


Fig. 5. Schematic view of a single detector module used in the "strip" PET. Position where the gamma quantum interacted can be determined from the difference between times measured at both edges of the strip.

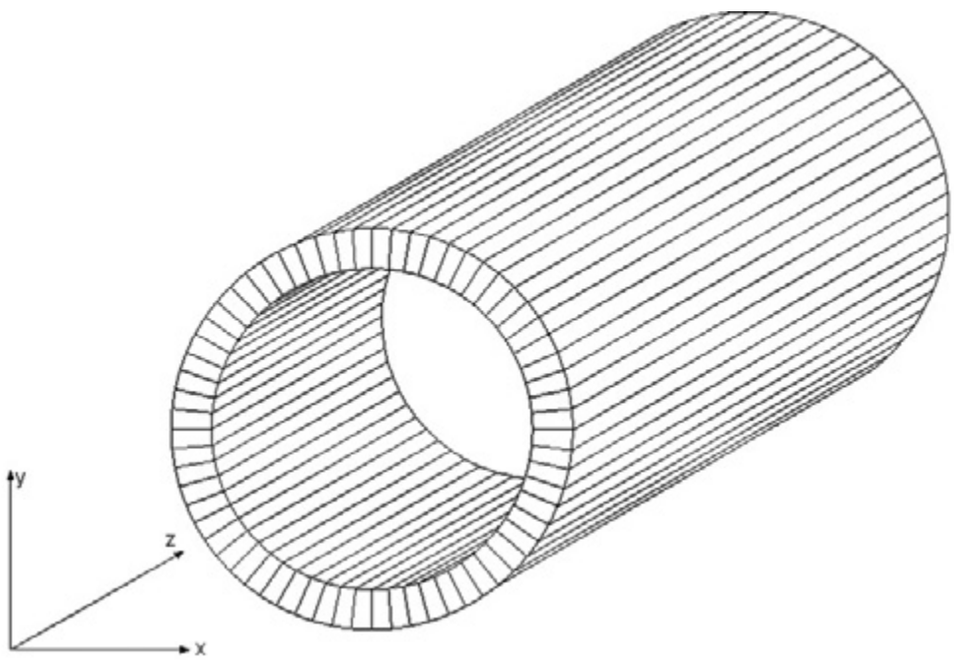


Fig. 6. Diagnostic chamber. Cylinder build out of scintillator strips.

The point of impact of a gamma quantum in a plane perpendicular to the axis of the strips can be determined from the position of a module which registered the signal, while the position along the scintillation chamber is determined using the difference between times measured in the front and rear photomultiplier. Energy of the electron colliding with gamma quantum is measured based on the amplitude of signals in the photomultipliers on both sides. Coincident registration of two gamma quanta allows to determine the line of response based on coordinates of reaction points reconstructed in both strips. The time of the reaction in each strip allows also to determine the annihilation point along the LOI based on the TOF method. The set of reconstructed LOR lines together with points of annihilation provides a tomographic picture.

Novel solution: Matrix PET

The Matrix-PET scanner would consist of organic scintillator plates, instead of currently used blocks of inorganic crystals. The plates could be set in many ways so as to cover the whole body of the patient, for example as it is shown in Fig. 7. The measurement of time and amplitude of light signals is carried out by photomultipliers matrix arranged around the chamber.

The interaction point within the plane of the plate can be reconstructed based on both: (i) the distribution of the time of the signals from the side photomultipliers and (ii) distribution of amplitudes of the recorded signals. Such solution allows also to determine the depth at which the gamma quantum has been

absorbed (DOI) on the basis of the distribution amplitudes of the signals from photomultipliers arranged on the sides (see Fig. 8 right). This feature allows to use thick plates without worsening of spatial resolution due to "the DOI problem" occurring in the current PET tomographs. Enlargement of the thickness enables efficient detection of gamma quanta using organic plastic scintillators, which are characterized by excellent time resolution, which is order of magnitude better in comparison with the fastest inorganic scintillators. Simultaneous registration of signals in blocks mounted in front of each other could enable determination of the gamma quantum path of flight. The time resolution through the use of plastic scintillators and as a result of simultaneous registration of the light signal by many photomultipliers would be much better than that achieved in the past PET scanners. Therefore, this solution would enable the usage of the TOF method permitting the determination of the annihilation point along the line of flight of annihilation quanta based on the time difference in reaching the different scintillation plates by the gamma quanta. Achieving TOF resolution of 50 picoseconds, possible when using organic scintillators and simultaneous measurement by many photomultipliers, would significantly simplify and hence make faster the reconstruction of the image.

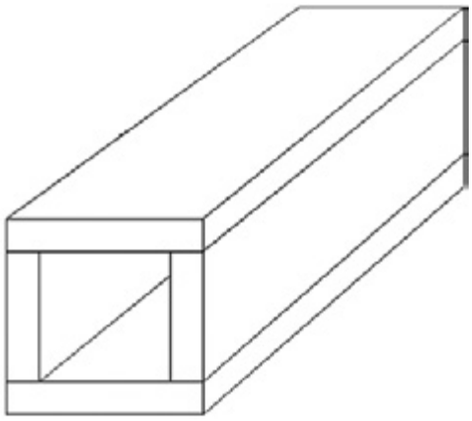


Fig. 7. Schematic view of the Matrix PET. Diagnostic chamber build out of scintillator plates.

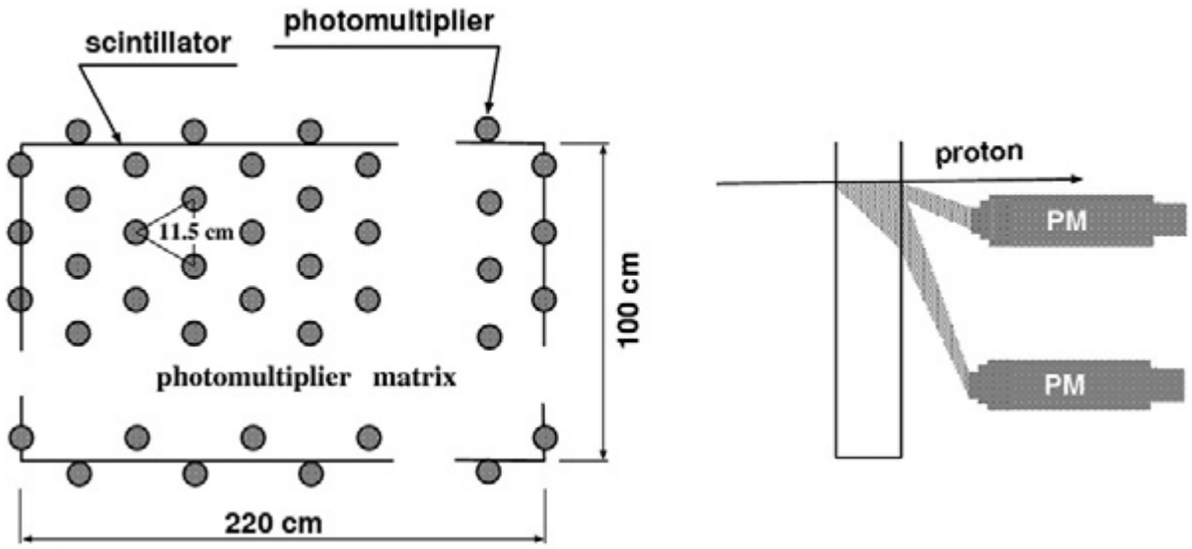


Fig. 8. Left panel presents schematic view of the scintillation wall detector as used in the particle physics experiment COSY-11 [12]; the panel demonstrates photomultiplier matrix together with the dimensions of scintillator wall. Right panel shows an edge of the scintillation chamber proposed for the Positron Emission Tomograph [2].

Summary

Both solutions Strip-PET and Matrix-PET offer an opportunity to obtain high time resolution and high acceptance, which are unaffordable in the current devices and which may translate into the precision of the annihilation point determination.

Proposed technology of construction of the PET tomograph requires the use of organic scintillators (plastics), which are relatively easy to produce in different shapes and large sizes, in contrast to currently used inorganic crystals. An ability of relatively easy extension of the size of the diagnostic chamber, is applicable particularly in "strip" PET where such extension does not entail an increase in the number of photomultipliers. This feature would decrease the construction costs of the PET scanners, which would also enable simultaneous imaging of the physiological processes throughout the whole body of the patient. Signals in currently used inorganic scintillators are much slower than the signals from the organic scintillators. Therefore, using fast organic scintillators (plastics) will also reduce the coincidence window giving the possibility to reduce accidental coincidences.

Acknowledgements

The authors are thankful to Dr. Gabriela Konopka Cupiał and the team of the Centre for Innovation, Technology Transfer and

University Development (CITTRU) for support in the investigations and for looking for companies and institutions interested in technology development, its testing and application. The work is also partly supported by The European Social Fund and Poland's Ministry of Treasury through the Scholarship for PhD Students „DOCTUS" and the „Małopolska Scholarship for PhD Students".

References

1. Moskal P.: Strip device and the method for the determination of the place and response time of the gamma quanta and the application of the device for the Positron Emission Tomography. Patent Application No. P 388 555 [WIPO ST 10/C PL388555], 2009.
2. Moskal P.: Matrix device and the method for the determination of the place and response time of the gamma quanta and the application of the device for the Positron Emission Tomography. Patent Application No. P 388 556 [WIPO ST 10/C PL388556], 2009.
3. Brauksiepe S. et al.: COSY-11, an internal experimental facility for threshold measurements. *Nucl. Instr. and Meth.* 1996, A 376: 397-410.
4. R. Czyżykiewicz et al.: Mechanism of the close-to--threshold production of the eta meson. *Phys. Rev. Lett.* 2007, 98: 122003.
5. OIL Kielce: http://www.nil.org.pl/xml/oil/oil56/gazeta/numery/n2007/n200705/ n20070506.
6. Jaworski Z.: Dobroczynne promieniowanie. *Wiedza i Życie* 1997, 3: 20-29.
7. Dobrzyński L.: Hormeza – Zjawiska powszechnie znane i nieznane, http://www.ipj.gov.pl/pl/szkolenia/matedu/hormeza.htm.
8. Zieliński M.: Feasibility study of the h'® π+ π- π0 decay using WASA-at-COSY apparatus, e-print arXiv: hep-ex/0807.0576.
9. Moses W.W.: Nuclear Medical Imaging-Techniques and Challenges, http://instrumentationcolloquium.lbl.gov/Nuclear%20Medical%20Imaging.pdf.
10. http://www.apace-science.com/proteus/lso.htm.
11. Conti M.: State of the art and challenges of time-of-flight PET. *Physica Medica* 2009, 25: 1-11.
12. Anton G. et al.: AMADEUS: A New type of large area scintillation detector with position, energy and time-of-flight determination. *Nucl. Instr. & Meth.* 1991, A310: 631-635.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 79-83

HUMAN SINGING AS A FORM OF BIO-COMMUNICATION

Edward Jacek Gorzelańczyk, Piotr Podlipniak

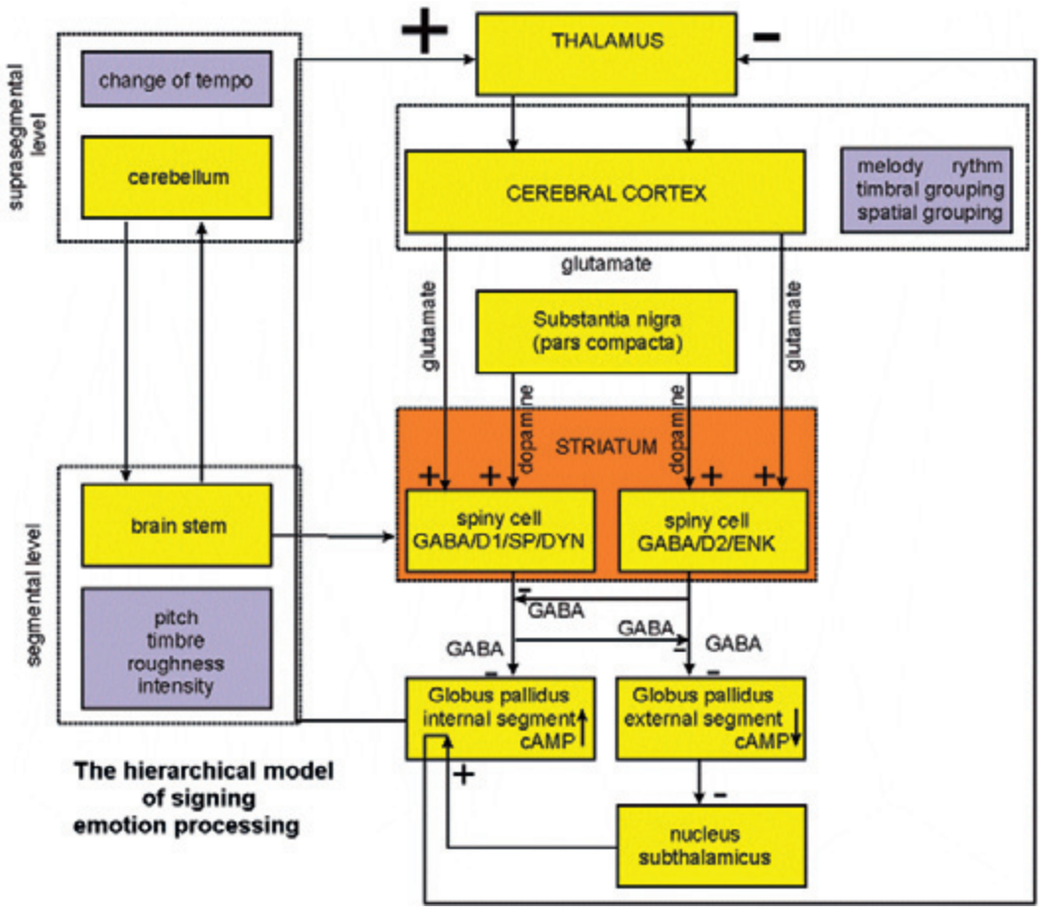
Nicolaus Copernicus University – Collegium Medicum, Bydgoszcz,
ul. Jagiellońska 13, Poland

Abstract: Most probably music, similarly to human speech, represents a biological adaptation [1], and singing is a mode of communication older than speech, present already in ancestors of *Homo sapiens* [2]. In various species of apes vocal expression has been demonstrated to be linked with expression of emotions [3], which indicates that singing is a carrier of emotional information which in evolution has appeared before formation of *Homo sapiens*. Hierarchical pattern of processing sound information in human cognitive system [4] allows to assume that singing may induce in the recipient both basic emotions and more complex reactions, linked to altered mood or induction of emotions. Processing of specific musical stimuli evokes specific emotional reactions [5]. Contemporary knowledge on processing of music in the nervous system and evolutionary perspective permit to distinguish such traits of musical course which code data on the type and intensity of emotions.
According to the authors, qualitative coding of principal emotions in musical course involves mainly segmental level using physical traits of the sound, such as intensity and timbre of sound while quantitative coding at the suprasegmental level involves mainly changes in tempo and intensity of sounds. In emotional communication conducted through a musical course the shared by the broadcaster and recipient set of culture-specific data on traits of music necessary for its correct processing in specific structures of nervous system, linked to cognitive processes, also plays a significant role.
In the study a hierarchical model of singing structure was suggested, which attempts to explain the way in which expression is coded and emotions are perceived in interpersonal communication.
Keywords: singing, vocal expression, speech, emotion, communication, coding of emotion, cognitive system

Introduction

Singing is a primitive way of music realization previous to any kind of instrumental music [6], probably older than speech [2]. But the ability of communicating by means of sounds is considerably evolutionary older than singing and speech [7]. The observations of contemporary living primate and other mammals indicate that these means of sound communication are partly shared [3, 8]. It is a result of the fact that in the brain evolution separate layers appeared, with distinct functions and over-growing the older structures but not replacing them, the older functions have been preserved [9, 10]. The phylogenetically newer brain structures (for example the prefrontal cortex) control evolutionary older structures connected to motor, emotional and cognitive functions [11].

In this perspective singing contains different degrees of motor, emotional and cognitive elements which are universal for the separate taxa (*Mammalia* – class, *Primates* – order, *Hominidae* – family, *Homo* – genus, *Homo sapiens* – species). In fact some of special forms of the human sound expression allow communication with other species (e.g. so called the pet – directed speech) [12]. Also many *Homo sapiens* behaviors in answer to sound stimuli show similarity to the animal's reactions. On the other hand, the exceptionality of the human musical behaviors and adaptive character of music suggest existence of the set of specific sound features which are exclusive to *Homo sapiens* and independent of culture [13]. One of these are the tonality and isometry which are music features strongly connected with emotion expression and which emerged as a result of natural selection. These features are absent in other kinds of the human sound communication. Of course a lot (perhaps majority) of musical information are specific to particular culture. Humans use all of these ways of communication in singing.



The singing and emotions

The human transmits various information such as spatial location, the structure of the body, sexual attractiveness, emotional states, cohesion of the group by means of sound carriers [14]. Some of them are present in all sound messages, other are connected only to some specific ways of man's sound expression. The singing is apart from speech one of more interesting forms of the man's sound expression. Obviously, the speech transfers mainly the semantic information whereas information carried by singing is more difficult to characterize. According to Darwin's opinion,"The human musical faculty must be ranked amongst the most mysterious with which he is endowed" [15]. In the western musical tradition, singing is understood as a form of art constructed by particular community in arbitrary way. However the basic function of singing is emotional communication between members of human community.

Philosophers, composers and musicologists affirmed many times that emotions are strongly related to music. Nevertheless, music was frequently assumed to express or represent emotions only [16] increasing evidence indicates that music represents a mode of direct emotional communication [17]. It was found that the emotions are activated in the brain in almost every music processing stage [18].The specific musical course can cause the changes in heart beat frequency and breath rate which indicates autonomic nervous system activation. Such physiological phenomena as well as concurrent motor and cognitive reactions [19] strongly influence emotion coding during singing.

In addition, the singer's emotional experience activates similar emotions at recipients better than instrumental realization of musical piece. It happens because the physiological changes of particular emotions (e.g. changes of breath rate dependent on fear level) impact on the voice feature [14, 19]. The close relationship between singing and emotions is probably a result of music adaptive function [1, 2]. The adaptive function of music represents a field of considerable discussion in the literature. For example Darwin proposed that music like bird's songs is important in the human mating [15]. In accordance to this conception the singing provide information of singer's attractiveness.

A second possibility for the adaptive function of music concerns its important role in increasing group integrity and consolidation [20, 21, 22]. Another adaptive function of singing derives from its crucial role in the care giving [23, 24, 25]. The above described phenomena have adaptive character because connected to them emotions make it possible to evaluate the biological importance of stimuli and to release the particular pattern of behaviors. If singing would have to inform about sexual attractiveness the particular features of song performances claim performer's fitness [26] and induce specific emotional reaction at listeners. In the second case, singing can consolidate group, for example by means of rhythmical synchronization or keeping tonal relations which evoked positive emotional reactions. In turn listening to a well synchronized loud singing of enemy's group evokes fear and awe. This function is present in the tribal warrior songs. The direct relationship between particular features of singing and babies' emotional reaction is observed during presentation of

lullabies [27]. It is possible that particular biological functions of music reach one to another and in different degree determinate the character of emotional expression in singing.

The evolutionary old layer of sound communication existing in speech and other spontaneous forms of sound communication plays an important role in arousing emotion by means of singing. Additionally, due to the common evolutionary sources of various forms of sound communication the emotional expression in singing has a hierarchic order. In this hierarchical system the most basic and effective in evoking emotional reaction are probably evolutionary older elements and the most subtle are these which were acquired in process of socialization and are conventional. Therefore primitive affective vocalizations are considered as information more authentic and are felt more truly than conventional and ritual information [19].

The evolution of emotional communication by means of sounds

The emergence and evolution of emotions is related to movement activity control [11]. The emotions make possible a selection of particular motor response in real environmental situation. Additionally the emotional communication allows recognizing emotions of different individuals and presenting one's own emotions. Therefore emotional communication is also adaptive. The emotional reaction to sound stimuli is universal among vertebrates and developed particularly among mammals. The elementary, evolutionary old and short-lasting affective states (for example the reaction of fright, disgust, pain, hunger) are processed by relatively simple neuronal circuits, which are localized on the lower parts of the central nervous system (for example on brain steam) [28]. The evolutionary newer emotions such as fear, anger, sadness, happiness are related mainly to limbic areas and trigger affective and cognitive reactions. These emotions have the common neuronal basis and generally do not show the significant interspecies differences among mammals [29, 30]. For the evolutionary newest emotional phenomena (sometimes called feelings) such as shame, feeling of guilt, remorse, empathy, romantic love, pride, the activation of cortical areas is necessary and essential [31].

The emotions can be activated and modified in various range by simple and/or complex sound stimuli [19, 35]. The emotional evaluation of sounds which come from environment and are connected with situations essential for survival (for example the sounds of attack and escape) has an adaptive value. The short, sudden, loud and low sound from loudspeakers arouses the reflexive reaction of surprise strong enough that it is difficult to inhibit this reaction even if the situation is not danger (for example during movie show in cinema). Also, the singer's emotions impact on his or her vocal apparatus functions in such way that listeners can recognize emotional state of a performer. In the case of high intensity of emotion the concurrent changes of muscles tension (for example of larynx muscles) as well as the increase of pressure in respiratory tracts directly and strongly influence produced sounds feature. It makes possible to evaluate the emotional singer's state.

Some of the emotions (for example sadness, anger, happiness) affect vocal expression significantly [14, 19]. Each of these basic emotions has a characteristic vocal acoustic signature [34]. Also the easiness of particular emotion expression differs from each other, which has adaptive reasons. For example the vocal expression of fear and anxiety is particularly clear in *Homo sapiens* species. The distinctive vocal expression of fear is adaptive because it facilitates and accelerates community members' reaction to approaching danger. On the contrary disgust is an emotion which is difficult to express by means of sounds [19].

The close and evolutionary old relationship between motor and emotion functions in vocal expression shows that the expression of emotion in singing has a motor nature. It was found that the even seemingly passive perception of music generates activation of motor cortex [32]. Also the intensity of emotion expression by means of sounds and gestures has the same direction (for example the pace of gesticulation and pace of musical course) [36] which additionally prove motor nature of emotion. It shows that the relationship between motor nature of singing and some metaphorical descriptions of musical phenomena related to rate, space and emotions (for example fast-slow music, high – low pitch of sound, violent- gentle melody) has not only cultural but also neurobiological and evolutionary reasons.

The hierarchic structure of emotional communication in singing

Singing like speech has two levels of phonological organization: segmental and suprasegmental [33]. The single tone at the segmental level in music is analogous to the phoneme in speech. The relative pitches of sound and relative time of duration are distinctive features of these smallest musical elements (musical phonemes). Obviously different sensational properties such as the timbre or loudness exist also on this level of music organization, however their modification does not change melody prototype. The timbre and loudness have however a great influence on the content of emotional meaning [34]. At the suprasegmental level individual units are put together to melodies.

In our opinion the way of emotional information coding in singing is compatible with the sequence of processing in human's nervous system. Hierarchical pattern of processing sound information in human cognitive system [4] allows assuming that singing may induce in the recipient both basic emotions and more complex reactions, linked to altered mood or induction of emotions. The basic and evolutionary old function of emotion is making a motor decision by organism according to biological meaning of environmental stimulus [11] (for example: avoidance-attraction or attack-escape behavior). Therefore the basic attribute of emotion is a valence. This kind of information is coded at the segmental level. Apart from valence at the segmental level the information about kind of emotion are coded. This information is attributed to sound of specific profile. Single silent and low sounds are usually connected with sadness. The emotional information coded in singing trigger affective response already at the beginning of processing, which are in turn controlled by evolutionary younger structures. But if the particular pattern of affective answer on sound emotional stimulus is evolutionary old and strongly preserved in the subcortical structures, it is difficult to inhibit it by brain cortex [11].

The intensity is another important characteristic of emotion. Also in singing the information about intensity of emotion is transmitted [35]. It is interesting that the singer can easier manipulate of emotional intensity expression than suppress particular emotion. This has a neurobiological reason. When the neuronal circuit processing of particular emotion is already active the full cognitive control of affection is not possible. The coding of the intensity changes of emotion is possible at suprasegmental level and consists in production of sounds with distinct parameters in time as well as their frequency. The most effective in the coding of the intensity changes of emotion is the use of dynamic changes and changes of musical tempo. The increase of emotion intensity is well coded by crescendo and accelerando, while decrease by decrescendo and diminuendo. Like in singing, the coding of emotion intensity is possible in speech as well as in prelingual systems of sound communication. This type of the emotion coding is present at various mammals. The emotional meaning of animal calls is comparable to specific dimension of human music, namely its expressive dynamics [8]. The express dynamics are best modeled by continuous variables in sharp contrast to the coding of emotional categories.

The majority of ways of emotion coding in singing described above are independent of culture. Of course, apart from these universal features of emotional expression the possession of culturally specific information is necessary to communicate emotions by means of singing [34]. The cultural information induces emotions and simultaneously activates cognitive processes in various degrees. The listener's cognitive engagement depends both on the kind of cultural information and his or her experience. Both universal and cultural features of singing induce emotions in a distinct way. However, because in the particular musical course these features are inseparable, it is difficult to estimate the qualitative and quantitative importance of them in the emotional communication.

Conclusion

The understanding of the way in which emotions are coded in singing may be helpful both in musical didactic work (indicating which of the performance elements our attention should be focused on in order to obtain desirable effects in listeners) as well as in the analysis and interpretation of music, which lacks until now any objective tools for the evaluation of emotional content of music. Studies on emotional expression in singing require that knowledge resources are applied from the area of humanistic arts (theory of music, history and esthetics of music, musical analysis, musical semiotics, cultural anthropology) as well as from natural sciences (biology, psychology, neuropsychology). Technical progress in imaging of cerebral functions and growing knowledge on functions of nervous system structures linked to human cognitive functions open potential for employing advances in natural sciences in musicology and to verify the till now accepted views.

References

1. Gray P.M., Krause B., Atema J., Payne R., Krumhansl C., Baptista L.: The Music of Nature and the Nature of Music. *Science* 2001, 291: 52-54.
2. Mithen S.: The Singing Neanderthals. The Origin of Music, Language, Mind, and Body. Harvard University Press: Cambridge 2006.
3. Hauser M.D.: The Sound and the Fury: Primate Vocalizations as Reflections of Emotion and Thought. In: Wallin N.L., Merker B., Brown S. (eds), The Origins of Music. MIT Press: London 1999, pp. 77-102.
4. Koelsch S., Siebel W.A.: Towards a neural basis of music perception. *Trends in Cognitive Sciences* 2005, 9/12: 578--584.
5. Juslin P.N.: Communicating Emotion in Music Performance: a Review and Theoretical Framework. In: Juslin P.N., Sloboda J.A. (eds), *Music and Emotion. Theory and research*. Oxford University Press: Oxford 2001, 309-337.
6. Geissmann T.: Gibbon Songs and Human Music from an Evolutionary Perspective. In: Wallin N.L., Merker B., Brown S. (eds), The Origins of Music. MIT Press: Cambridge–London 2000, pp. 103-123.
7. Tembrock G.: Tierstimmenforschung. Eine Einführung in die Bioakustik. Ziemsen Verlag: Wittenberg 1977.
8. Merker B.: Is There a Biology of Music, and Why Does it Matter? In: Kopiez R., Lehmann A.C., Wolther I., Wolf Ch. (eds), Proceedings of the 5th Triennial ESCOM Conference. *Hanover University of Music and Drama* 2003, 402-405.
9. Roederer J.: On the Concept of Information and Its Role in Nature. *Entropy* 2003, 5/1: 3-33.
10. Striedter G.F.: Principles of Brain Evolution. Sinauer Associates, Inc. Publishers: Sunderland 2005.
11. Gorzelańczyk E.J., Włodarczyk K., Laskowska E., Kunc M., Jelonek J.: Digitalized Drawing Test (Ddt) – The Tool For An Early Subclinical Motor Symptoms Diagnosis. *Bio-Algorithms and Med-Systems* 2008, 4/8: 93-99.
12. Stockmann D.: Musik – Sprache – Tierkommunikation. *International Review of the Aesthetics Sociology of Music* 1979*,* 10: 5-45.
13. Podlipniak P.: Uniwersalia muzyczne. Wydawnictwo Poznańskiego Towarzystwa Przyjaciół Nauk: Poznań 2007.
14. Karpf A.: The Human Voice. Bloomsbury Publishing: New York–London 2006.
15. Darwin Ch.R.: The descent of man, and selection in relation to sex. Vol. II. John Murray: London 1871.
16. Meyer L.B.: Emotion and Meaning in Music. The University of Chicago Press: Chicago–London 1956.
17. Krumhansl C.: An Exploratory Study of Musical Emotions and Psychophysiology. *Canadian Journal of Experimental Psychology* 1997, 51/4: 336-353.
18. Panksepp J., Bernatzky G.: Emotional Sounds and the Brain: the Neuro-Affective Foundations of Musical Appreciation*. Behavioral Processes* 2002, 60: 133-155.

19. Johnstone T., Scherer K.R.: Vocal Communication of Emotion. In: Lewis M., Haviland-Jones J.M. (eds), Handbook of Emotions. The Guilford Press: New York 2000, pp. 220--235.
20. Roederer J.: The Search for a Survival Value of Music. *Music Perception* 1984, 1: 350-356.
21. Ridley M.: The Origins of Virtue: Human Instincts and the Evolution of Cooperation. Penguin Books Ltd.: London 1997.
22. Hagen E.H., Bryant G.A.: Music and Dance as a Coalition Signaling System. *Human Nature* 2003, 14/1: 21-51.
23. Papousek H.: Musicality in Infancy Research: Biological and Cultural Origins of Early Musicality. In: Deliege I., Sloboda J. (eds), Musical Beginnings. Oxford University Press: Oxford 1996, pp. 37-55.
24. Dissanayake E.: Root, Leaf, Blossom, or Bole: Concerning the Origin and Adaptive Function of Music. In: Malloch, S., Trevarthen, C. (eds), Communicative Musicality. Oxford University Press: Oxford 2009, pp. 17-30.
25. Falk D.: Finding Our Tongues: Mothers, Infants and the Origins of Language. Basic Books: New York 2009.
26. Miller G.: Evolution of Human Music through Sexual Selection. In: Wallin N.L., Merker B., Brown S. (eds), The Origins of Music. MIT Press: London 2000, 329-360.
27. Trehub S.E., Trainor L.: Singing to Infants: Lullabies and Play Songs. *Advances in Infancy Research* 1998, 12: 43--77.
28. Denton D.: The Primordial Emotions. The Dawning of Consciousness. Oxford University Press: New York 2005.
29. Ekman P.: Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. Times Books: New York 2003.
30. Pankseep J.: Affective Neuroscience: the Foundations of Human and Animal Emotions. Oxford University Press: New York 1998.
31. Damasio A.R.: Descartes Error. Emotion, Reason, and the Human Brain. Avon Books: New York 1994.
32. Janata P., Grafton S.T.: Swinging in the Brain: Shared Neural Substrates for Behaviors Related to Sequencing and Music. *Nature Neuroscience* 2003, 6/7: 682-687.
33. Bielawski L.: Muzyka jako system fonologiczny. *Res Facta* 1968, 3: 166-171.
34. Welch G.F.: Singing as Communication. In: Miell D., MacDonald R.; Hargreaves, D. (eds), Musical Communication. Oxford University Press: New York 2005.
35. Scherer K.R.: Expression of Emotion in Voice and Music. *Journal of Voice* 1995, 9(3): 235-248.
36. Clynes, M.: Sentics: The Touch of Emotions. Doubleday/ Anchor: New York 1977.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 85-93

PHOTODYNAMIC THERAPY OF MELANOMA. MONTE CARLO MODELING OF LIGHT TRANSPORT IN HUMAN PIGMENTED SKIN

ZENON MATUSZAK

Department of Medical Physics and Biophysics, Faculty of Physics and Applied Computer Sciences, AGH-University of Science and Technology, Al. Mickiewicza 30, 30-059 Cracow
matuszak@novell.ftj.agh.edu.pl

Abstract: Among many available methods used in clinical practice to treat cancer tissues, Photodynamic therapy (PDT) is a relatively new tool. PDT is an enhancement of chemotherapy. To destroy cancerous tissue, photosensitizer (PS), oxygen (O_2) and light – three crucial elements of PDT – have to work together to produce sufficiently high concentration of cytotoxic reagents, reactive species of oxygen (ROS), mainly singlet oxygen $^1O_2$. To effectively generate $^1O_2$, light intensity needed for the activation PS should achieve appropriate value in the place of the location of PS within the cancerous lesion. In pigmented cancers amount of melanin is larger than in normal tissues due to proliferation of melanoma cells. Melanin strongly absorbs light and attenuates its intensity. To investigate the influence of melanin on PDT efficiency, a seven-layer computer skin model was proposed and tentatively examined in this work. Influence of melanin on light intensity distribution was simulated with Monte Carlo method. For some fluence rate values (50,100,150 and 200 $mWcm^{-2}$) the computer simulations of PDT effect using parameters of porphyrin photosensitizer (tritiolyloporphyrin, λ_{ex}=650 nm) and kinetic schema of PDT based on Jablonski diagram were analyzed. Simulations have shown that fluence rate values strongly modify efficiency of therapy and in the case of low fluence rate values deeper laying cancerous cell would not be killed due to too low $^1O_2$ production. Additionally it was shown that the problem of estimation of melanin content in tissue can be solved by measurement of diffuse reflectance of skin.
The possibilities of further development of the model were also briefly discussed.
Keywords: photodynamic therapy, melanin, skin, light transport, Monte Carlo, modeling

Introduction

Skin cancers are a growing problem in the world. Amongst many kinds of skin cancer, cutaneous malignant melanoma (simply melanoma) is the most dangerous and its occurrence systematically increases. Melanoma detected at early stages of development (no metastasis) is curable close to 100% but precise diagnosis is very difficult because the early melanoma often resembles another pigmented lesions [1, 2]. The non-invasive methods which can provide quantitative, physically and physiologically meaningful measurements of tissue properties are an attractive tool for the study, diagnosis and treatment of these cancers. Among existing methods which could be used to this purposes, it seems that the most promising are the optical ones. The optical methods of diagnosis, such as direct visual inspection, dermoscopy together with epiluminescence microscopy are the methods which are used in everyday dermatological practice [3, 4] and some atlases of dermoscopic images which facilitate the diagnosis have been published [1]. All of the above mentioned methods are based on the spectral properties of melanin, because color of the pigmented skin lesions is used to identify the kind of the lesion and its spreading. This inspection allows to choose the area of therapy, in case of photodynamic therapy (PDT), area of lightening.

Photodynamic therapy is a kind of chemotherapy which uses the light to excite drugs, photosensitizers (PS) to produce Reactive Oxygen Species (ROS) which are factors destroying cancer cells. The effectiveness of ROS production strongly depends among others on the light intensity in the place of PS localization inside the tissue i.e. on optical properties of melanin and skin as a whole and of properties of the PS itself [5].

PDT is seen as a promising alternative in the treatment of melanomas, despite the strong absorbing properties of melanin in a UV-VIS range and its protecting properties against ROS [6]. Many of PS, absorbing in various spectral area, were tested against melanoma cells, the experiments have been shown

effectiveness of PDT procedure in killing melanoma cells [7, 8, 9, 10, 11].

PDT is a rather complicated kind of therapy, its success being related to availability in place of action of three elements light, PS and oxygen ($^3O_2$) which have to cooperate to produce the sufficient amount of $^1O_2$ within tissue to effectively kill cells. To quantify the PDT effectiveness the quantity called ROS threshold value (ROS_{th}), denoting the quantity of ROS species necessary to kill the cell, is introduced [12]. In a case of so called II-type PDT, ROS_{th} could be determined as appropriate $^1O_2$ concentration. There is the simple rule of thumb for PDT dosimetry [12] which takes into account all the essential components of therapy and specifies the depth of tissue necrosis and quantity generated ROS during PDT:

$$ROS_{th}=E_0\, k t \exp(-z_{necr}/\delta) C b \Phi_T\, \Phi_\Delta\, f_R \qquad (1),$$

where E_0 [W/cm^2] – irradiance treatment light onto the tissue surface; t[s] – exposure time for used light; δ [cm] optical penetration depth of used light; k [dimensionless] – augmentation of light at surface due to backscattering; z_{necr} [cm] – depth of necrosis zone, e [cm^{-1}/(mg/g)] – extinction coefficient of PS, C [mg/g] – concentration of PS (per g of tissue); b [ph/J] – conversion factor, b=λ/hc, where h is Planck constant, c – light velocity in vacuum [m/s], λ – wavelength in nm; Φ_T [dimensionless] – quantum efficiency for PS triplet state formation, Φ_Δ – quantum efficiency for $^1O_2$ formation, which depends on $^3O_2$ concentration; f_R –fraction of oxidizing species that attack critical cell sites; ROS_{th} [ph/cm^3] – threshold value of oxidizing species concentration leading to cell death.

All components in the eq (1) must be determined separately. The PS concentration is usually measured using fluorescence [13], while the electrochemical methods are used to determine $^3O_2$ [14]. Light fluence rate (photon concentration) can be calculated by solving Radiative Transfer Equation (RTE)

$$\frac{1}{v}\frac{\partial L(r,s,t)}{\partial t} = -s\cdot\nabla L(r,s,t) - (\mu_a+\mu_s)L(r,s,t) + \mu_s \int_{4\pi} L(r,s,t)p(s,s)\, d\Omega(s) + Q(r,s,t) \qquad (2),$$

where L is a radiance, r – position vector, s – direction vector, p(s',s) – scattering phase function, μ_a – absorption coefficient, μ_s- scattering coefficient, Q – source function, v – light velocity in a medium, [15]. Some attempts are being taken to determine the $^1O_2$ directly using its 1270 nm specific luminescence [16].

Many methods of different scope of applications were developed to solve RTE and to calculate the distribution of radiation field within tissues [17], among them special position occupies the Monte Carlo method, which have been used extensively in simulation light transports due to its simplicity and elasticity [18]. This method is also used in this work.

The main goal of this work was to investigate the influence of melanin content on the effectiveness PDT via its influence on fluence rate values. The work is divided into two parts, in the first part of the work the influence of melanin content on intensity distribution within normal and melanized skin represented by optical multi-layer model was simulated, in the second part kinetic behavior of PDT important species was investigated. Kinetic schema of PDT was analyzed based on Jablonski diagram describing the main PDT events. The calculations were made for one of PS belonging to porphyrines family, tritolyloporphyrin (TD) [9].

Materials and Methods

The PDT simulation model

In the II type of Photodynamic Therapy (II-type PDT) ROS_{th} concentration is identified with $^1O_2$ concentration which have to exceed a certain threshold dose in cell to kill it. The cell survival determines the PDT outcome. To determine $^1O_2$ concentration, the coupled rate equations describing interactions between the PS, $^3O_2$, intracellular targets and light fluence rate must be solved. In the considered model it is assumed that the initial concentrations of $^3O_2$ and PS are given, but to obtain fluence rate values inside skin the RTE must be solved for given optical properties of postulated tissue model.

Basic photophysical and photochemical processes involved in type II PDT are shown in Fig.1.The Jablonski (energy levels) diagram summarizes the formation and decay of $^1O_2$ produced during the II- type PDT [5, 19]. The created $^1O_2$ reacts with intracellular targets and these reactions cause the cell-destroying events. $^1O_2$ can also react with PS itself leading to self-destruction of PS (the so called photobleaching process) and could also be deactivated by radiation processes. Because of its high reactivity, $^1O_2$ has a short lifetime in tissue, usually less than 0.01 – 0.04 μs in a biological environment, and its diffusion distance is about 10-20 nm, which is important for the sake of patient's safety. PDT is local therapy with small incidental effects [5].

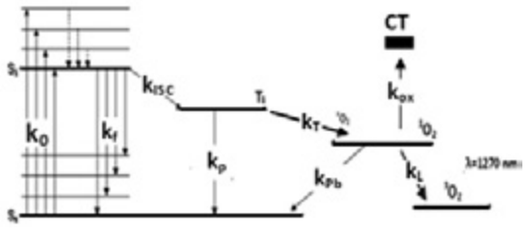


Fig. 1. Jablonski diagram of $^1O_2$ formation and decay during type-II PDT. S_0-ground state of PS, S_1- singlet excited state of PS, T_1- triplet state of PS, $^3O_2$- ground state of oxygen, $^1O_2$- singlet oxygen, CT-cellular targets, k_i – rate constants for processes considered in model; reactions form are given below. k_0-rate constant of excited singlet state of PS (S_1) formation from ground state of PS (S_0); k_f- fluorescence rate constant, k_{ISC}- rate constant of intersystem crossing process, k_p- phosphorescence rate constant, k_T-rate constant of $^1O_2$ formation (energy transfer from PS triplet state to $^3O_2$), k_{Pb}- photobleaching rate constant, k_L- rate constant of $^1O_2$ luminescence (λ=1270 nm), k_{ox}- rate constant for oxidation of cellular targets(CT).

The sequence of kinetic reactions associated with photochemical processes involved in II-type PDT is given below:

$$S_0 + h\nu_{ex} \xrightarrow{k_0} S_1 \quad (3)$$

$$S_0 \xrightarrow{k_f} S_1 + h\nu_f \quad (4)$$

$$S_1 \xrightarrow{k_{ISC}} T_1 \quad (5)$$

$$T_1 \xrightarrow{k_p} S_0 + h\nu_p \quad (6)$$

$$T_1 + {}^3O_2 \xrightarrow{k_T} S_0 + {}^1O_2 \quad (7)$$

$${}^1O_2 + S_0 \xrightarrow{k_{Pb}} {}^3O_2 + S_{ox} \quad (8)$$

$${}^1O_2 \xrightarrow{k_L} {}^3O_2 + h\nu_L \quad (9)$$

$${}^1O_2 + CT \xrightarrow{k_{ox}} {}^3O_2 + CT_{ox} \quad (10)$$

where $h\nu_e$, $h\nu_f$, $h\nu_p$, $h\nu_L$ are photon energies of exciting, fluorescent, phosphorescence and $^1O_2$ luminescence radiation respectively, h – is a Planck constant, v – frequency of radiation. S_{ox} and CT_{ox} are products of irreversible oxidation of PS and CT(cellular targets) by $^1O_2$, respectively. Other symbols in equations have the same meaning as in Fig.1. The rate constant k_0 (in Eq.3) is given as

$$k_0 = \sigma\phi \quad / \quad (h\nu_{ex}) \quad (11),$$

where σ [cm^2] is absorption cross section of PS ground state and $\sigma[cm^2]$=3.82 x 10^{-21} ε (12), where ε is extinction coefficient, Φ [Wcm^{-2}] – radiation fluence rate, $h\nu_{ex}$ – excitation photon energy (in J).

Based on reactions involved in II- type PDT (eqs.3-10) and using the mass action law, the rate equations system can be derived and used to analyze the intracellular photochemical processes in PDT [20]. Presented model is composed of six equations (six state variables, their concentrations are denoted by []) and has the following form:

$$\frac{d[S_0]}{dt} = -k_0[S_0] + k_f[S_1] - k_{Pb}[{}^3O_2][S_0] + k_p[T_1] + k_T[T_1][{}^3O_2] \quad (13)$$

$$\frac{d[S_1]}{dt} = k_0[S_0] - (k_f + k_{ISC})[S_1] \quad (14)$$

$$\frac{d[T_1]}{dt} = k_{ISC}[S_1] - k_p[T_1] - k_T[T_1][{}^3O_2] \quad (15)$$

$$\frac{d[{}^3O_2]}{dt} = -k_T[T_1][{}^3O_2] + k_L[{}^1O_2] \quad (16)$$

$$\frac{d{}^1O_2}{dt} = k_T[T_1][{}^3O_2] - k_L[{}^1O_2] + k_{ox}[CT][{}^1O_2] - k_{Pb}[{}^1O_2][S_0] \quad (17)$$

$$\frac{d[CT]}{dt} = -k_{ox}[CT][{}^1O_2] \quad (18)$$

The system of coupled ordinary differential equations (ODE) 13 to 18 which describes time behavior of all species engaged in PDT was solved using COPASI system [21]. COPASI (Complex Pathway Simulator) uses LSODA solver (Livermore Ordinary Differential Equations Solver) to solve systems of ODE, including stiff systems. Some additional calculations were performed using Matlab's (Mathworks Inc) ordinary differential equations solver ODE15S for stiff systems of ODE.

In the model the reaction rates k_i remain constant for a given kind of PS depending on the photochemical properties of PS, but k_0 in this model must be calculated. The k_0 value at position r inside tissue depends on fluence rate value (eq.3, 11) at that position for given excitation wavelength, which is a function of tissue optical properties. The calculation of fluence rate values within melanized tissue requires solving of the light transport problem in such tissues.

Light transport trough melanized tissue

Theoretical modeling of light propagation in biological media is usually based on Radiation Transport Theory (RTT) [15]. In frames of RTT, optical properties of tissue are described by parameters characterizing interaction of photons with tissue structures: absorption and scattering coefficients and phase function. These parameters are functions of wavelengths [22].

Optical properties of skin layers

Generally skin optics is very complicated by the facts that skin is an irregularly shaped, inhomogenous, multi-layered structure and has anisotropic physical properties [24, 25, 26]. As shown in Fig. 2, skin represents a complex heterogenous medium, where spatial distribution of main chromophores, blood and melanin are variable with depth. The optical properties of skin components on each level differ significantly depending on their physico-anatomical characteristics. It is possible to define the separate anatomical regions where the skin cell structures, chromophore content and blood concentration are roughly constant. This fact allows us to approximate skin as a multi-layered medium containing few layers. Each homogeneous layer has spectral absorption and scattering coefficients depending on its chemical and physical structure. This approach is universally used and skin model considered in this work is similar to previously described in literature, spectral dependence of optical parameters for individual skin layers as is described below are taken from various experiments conducted not only on human skin, but also on rat skin and skins of other animals [24, 25, 26].

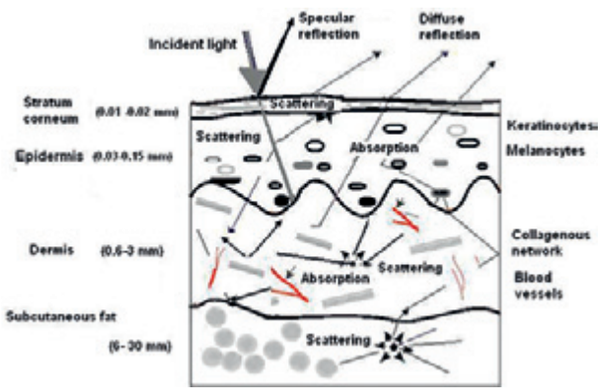


Fig. 2. Schematic diagram of skin in cross section showing the main optical interactions between light and skin components. The light passes through the thin stratum corneum into epidermis. The epidermis propagates and absorbs light, absorption comes from melanin, the scattering in the forward direction towards the dermis (Mie scattering) is caused predominantly by cell membranes and nuclei. Next light propagates into dermis,

where light is multiply scattered by collagen fibers (Mie scattering) and cell organelles (Rayleigh scattering). In dermis the light absorption is due to hemoglobin presence. Subcutaneous fat only scatters light.

The outermost section of skin is stratum corneum approximately thick 0.01-0.02 mm. It is often considered a part of epidermis and it is composed from dead cells [25, 27, 29]. Light absorption in this layer is very low. The top layer epidermis (sometimes called epithelium), thick 0.027- 0.15 mm is formally composed of five layers [25, 28, 29]. It is the most interesting layer from our point of view, because melanin is produced in this layer, in its deepest sub-layer, so-called stratum germinativum (stratum basale). Epidermis does not contain any blood vessels. Spectral properties of epidermis in UV-VIS depend on its melanin content. There are two types of melanins brown/black eumelanin and red/brown pheomelanin [25, 28]. Skin color is predominantly associated with presence of eumelanin (pheo-/eumelanin ratio is about 1/8, but varies from individual to individual) and above all depends on the fraction of epidermis occupied by melanosomes, organelles containing melanin. That fraction varies from 1% to 3% in Caucasians but in dark Africans it goes up to 43% [25, 28].

The lower skin layer, dermis, is a 0.6-3 mm thick structure divided into two sub-layers: the papillary dermis and the reticular dermis [25, 27, 29]. The dermis is mostly comprised of collagen (70%), vessels and fatty tissue. Absorption in this layer is dominated by a natural chromophore of blood cells hemoglobin, which appears in two forms: oxygenated (in arteries 90%, in veins about 47%) and deoxygenated. Each of these forms has a little different absorption spectrum. The volume fraction of blood in tissue changes from 0.2% to 7% [25, 29]. The last layer is the hypodermis, subcutaneous tissue which is not considered a part of skin. It can be up to 3 cm thick. It is built form white fat. Absorption in this layer is negligible [25, 29].

Melanoma begins with a proliferation of alternated melanocytes, which may penetrate the basement membrane of the epidermis and invade the deeper skin layers, dermis and at the end, the subcutis [1]. In metastasis, melanoma spreads to remote sites in the organism. The development of melanoma changes significantly the distribution of the pigment in skin layers and this fact must be taken into account in the model of light propagation through appropriate modification of the absorption properties of invaded skin layers. The process of creating changes of pigment distribution within the skin, characteristic for the progress of the pigmented melanoma, is pictured in Fig. 3.

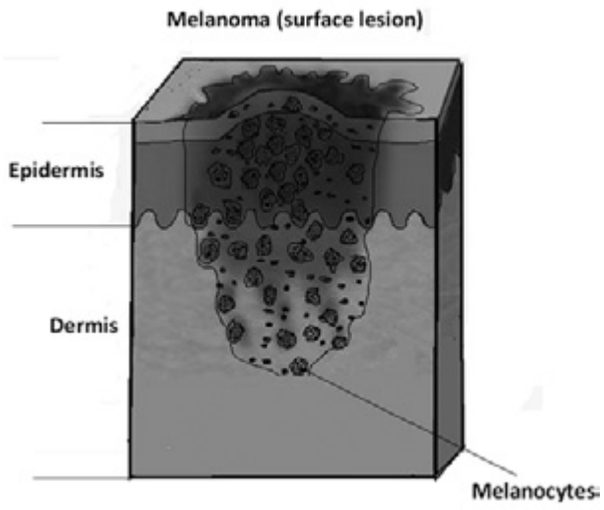


Fig. 3. Skin melanoma development from normal skin. Dividing melanocytes migrates from the epidermis into other skin layers. The proliferation of melanotic cells changes the optical properties of neighboring skin layers increasing their absorption.

Optical model of skin

To simulate the interaction of light and skin, optical parameters of skin layers must be calculated. Each of the optically homogeneous layer has its own spectral absorption and scattering properties depending on its structure and composition. As mentioned above, the data necessary to perform simulations of light transport in tissue are collected from literature [23, 24, 25, 26] and the presented model is similar to previously reported in literature, especially in the computer graphics, in modeling skin reflectance and skin rendering [29, 30, 31, 32]. Optical characteristics of skin layers in considered model are given below.

The stratum corneum is a low absorbing skin layer. Its absorption changes when it is invaded by melanocytes during melanoma development. There are no data about content of melanosomes (or melanin itself) in this layer during invasion but a visual inspection allows as to assume that it is no more than 8%.The absorption coefficient of stratum corneum can be estimated as:

$$\mu_a^{strcor}(\lambda)=(f_{skinb})(\mu_a^{skinb}(\lambda))+(1-f_{skinb})(\mu_a^{mel}(\lambda)) \quad [cm^{-1}] \quad (19),$$

where $\mu_a^{skinb}(\lambda)$ is an absorption coefficient due to presence of absorbers other than melanin and hemoglobin, $\mu_a^{mel}(\lambda)$ is absorption coefficient of melanosomes, f^{skinb} is volume fraction of these absorbers in layer. Spectral dependence of $\mu_a^{skinb}(\lambda)$ and $\mu_a^{mel}(\lambda)$ are given below.

The epidermis absorbs light depending on amount and type of melanin in the skin. Absorption coefficient for epidermis is assumed as:

$$\mu_a^{epid}(\lambda)=(f_{mel})(\mu_a^{mel}(\lambda))+(1-f_{mel})(\mu_a^{skinb}(\lambda)) \quad [cm^{-1}] \quad (20),$$

where $\mu_a^{mel}(\lambda)$ is an absorption coefficient of melanin, f_{mel} is the total volume fraction of melanin in the epidermis, $\mu_a^{skinb}(\lambda)$- ab-

sorption due to other components, λ - the light wavelength. The spectral dependence of μ_a^{mel} (λ) and μ_a^{skinb} (λ) are described by power laws:

$$\mu_a^{mel}(\lambda)=(6.6\times10^{11})(\lambda^{(-3.33)})\ [cm^{-1}] \quad (21),$$

$$\mu_a^{skinb}(\lambda)=(7.84\times10^{8})(\lambda^{(-3.255)})\ [cm^{-1}] \quad (22)\ [25, 26, 30, 31].$$

The baseline absorption approximation formula is based on measurements of bloodless rat skin (omitting the hemoglobin and melanin absorption) and it was assumed that it has the same form for both epidermis and dermis [25, 26], in this work is assumed that this formula applies also to stratum corneum.

The absorption coefficient for dermis combines similar to epidermis the baseline absorption coefficient and a blood absorption coefficient and is given by formula:

$$\mu_a^{der}(\lambda)=(f_{blood})(\mu_a^{blood}(\lambda))+(1-f_{blood})(\mu_a^{skinb}(\lambda)) \quad [[cm^{-1}] \quad (23),$$

where f_{blood} is an average volume fraction of blood over the whole body, which is not the case in reality, particularly in melanomas. Typical value of f_{blood} 0.2% , in skin it is likely to be about 2-5% [17]. The principal pigment of blood is hemoglobin and its normal concentration in blood is 150g/liter. In case of melanoma this expression is modified by including an additional element including presence of melanocytes in dermis as it is suggested in Fig. 3.

$$\mu_a^{der}(\lambda)=(f_{blood})(\mu_a^{blood}(\lambda))+(1-f_{blood})(f_{mel})(\mu_a^{mel}(\lambda))+(1-f_{blood})(1-f_{mel})(\mu_a^{skinb}(\lambda)) \quad [[cm^{-1}] \quad (24),$$

The $\mu_a^{blood}(\gamma)$ has two components connecting to two forms of hemoglobin oxy- and deoxyhemolgobin (HbO_2 and Hb) which have slightly different absorption bands. Fraction of hemoglobin in blood γ is connected to blood hematocrit Hγ , where is the volume fraction of erythrocytes in total volume of all blood cells, and is the volume fraction of hemoglobin in erythrocytes. This difference between F_{Hb} and F_{RBC} is omitted because in the considered wavelength region the differences in absorption between two forms of hemoglobin are negligible. Water absorption is only important in NIR (Near Infrared) area of wavelength [35] and its participation is omitted here.

Scattering properties of skin are modeled by a combination of the Mie and Rayleigh theory. The total skin scattering coefficient as $\mu_{scat}^{tot}(\lambda)$ a function of wavelength λ is described by the following formula:

$$\mu_{scat}^{tot}(\lambda)=\mu_{scat}^{Mie}(\lambda)+\mu_{scat}^{Ray}(\lambda)=(2\times10^{5})\ \lambda^{-1.5}+(2\times10^{12})\ \lambda^{-4}\ cm^{-1} \quad (26),$$

the scattering coefficient, where $\mu_{scat}^{Mie}(\lambda)$ describes Mie scattering behavior and $\mu_{scat}^{Ray}(\lambda)$ the Rayleigh scattering. From experiments it is known that Henyey – Gernstein (HGPF) phase function p_{HG} ($\cos\theta$) fits well the scattering behavior of epidermis and dermis.

$$p_{HG}(\cos\theta)=\frac{(1-g^2)}{2(1+g^2-2g\cos\theta)^{3/2}} \quad (27),$$

where θ is a scattering angle [25]. The reduced scattering coefficient is defined as $\mu_{scat}'(\lambda)=\mu_s(\lambda)(1-g)$ (28), where g is an anisotropy factor (mean cosine of deflection angle) and it takes in the range g= 0.7-0.95 [25]. Backward scattering in epidermis is week, for the wavelength dependence anisotropy factors for epidermis (g_{epi}) and dermis (g_{der}) was assumed as $g_{epig} \cong g_{der}$ $0.62+0.29\times10^{-3}\ \lambda$ (29) [25, 26].

The skin model considered in this work consists of seven layers. Values of optical parameters for each of considered skin layers in wavelength were calculated for several wavelength values from the scope between 620 – 670 nm. The optical parameters of skin layers used in fluence rate simulations for normal skin are shown in Table 1.

Layer	d x10^{-4}(cm)	μ_s x10(cm^{-1})	μ_a x10(cm^{-1})	g	n
Stratum corneum	20	100	0.8	0.9	1.5
Epidermis	80	60	0.15	0.85	1.34
Dermis I(papillary)	150	30	0.07	0.9	1.39
Dermis II(upper blood)	80	35	0.1	0.95	1.4
Dermis III (reticular))	1500	25	0.05	0.76	1.39
Dermis IV (deep blood)	80	35	0.15	0.95	1.4
Subcutenous fat	6000	15	0.075	0.8	1.44

Tab. 1. The optical parameters of the seven-layer human skin model. The parameters were used in the calculation of fluence rate distribution. d – thickness of the layer, μ_a – absorption coefficient, μ_s – scattering coefficient, g – anisotropy factor, n – refraction index. (λ=650 nm)nm. Data were calculated using equations (19-29). Same values were taken from literature [22, 25, 29, 31].

The progression of melanoma was taken into account through the change of absorption coefficient values of the skin layer, according to rise of the amount of melanin in the layer as is shown in Fig. 3.The simulations of light fluence rate distribution (eq.2) were performed using Monte Carlo method. In the simulation two public domain codes MCML and CONV written in C were used, details of the codes are described in [12, 32, 33]. Results were visualized using the Matlab. All calculation were done on PC computer (2GHz, dual core with 4Gb RAM).

Results

The essential purpose of the work was the theoretical examination of the modifying influence of the melanin content on effectiveness of PDT in the treatment of melanoma. The preliminary calculations of normal skin optical properties using presented skin model for chosen wavelengths between 620-670 nm have shown that their values practically differ to a little extend. Only change of melanin content in the skin layers due its absorbing

properties strongly modifies light distribution within the tissue, influences on receiving by light intensity the level necessary for effective cell killing. During the development of melanoma the amount of melanin containing cells in skin layers increases and therefore as a result strongly changes the optical properties of tissue. Dependence of absorption coefficient of epidermis on melanin content is shown in Fig. 4.

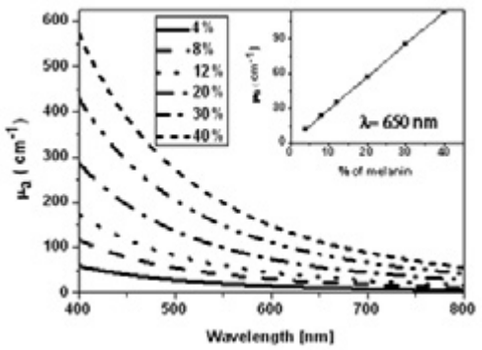


Fig. 4. Absorption coefficient of epidermis as a function of percentage content of melanin. Values calculated using formula (20). Inset. An absorption coefficient of epidermis at chosen wavelength (λ=650 nm) is a linear function of melanin content.

Strong melanin absorption suggest for wavelength below 600 nm suggests that potentially effective anti-melanoma PS should absorb above 600 nm where the absorption coefficient of melanin is relatively low. Such a PS should also have a relative high value of extinction coefficient in this wavelength area to be able to absorb light on large depths in the tissue to a considerable degree. The compounds that fulfill this conditions belong for example to the family of porphyrines or phtalocyanines. Both classes of PS were already tested as potentially promising dyes in PDT of melanomas [34, 35].

Influence of various melanin content on light fluence rate distribution in considered seven-layer skin model is shown in Fig. 4. and Fig. 5. Monte Carlo simulations of fluence rate distribution were performed using optical parameters (λ_{ex}=630, 650 nm) taken from Table 1. The progress of melanoma manifesting in darkening of tissue invaded by the cancer, was taken into account in simulations through the change of optical parameters (absorption coefficient) in the considered skin layers in Table 1. Fig. 5B shows the fluence rate distribution in normally melanized skin, Fig. 5A – the light distribution in pigmented melanoma. A strong influence of melanin content on light penetration depth (z- direction) and on lateral distribution of scattering radiation (r-direction) is observed. The presence of melanin decreases both, depth of light penetration and radial range of scattered light within tissue. The decrease of the light intensity is particularly strong in epidermis and in upper dermis. Comparing the simulation results for normal and melanized tissues (Fig. 5) it is noteworthy that in the case of considerable melanized tumor deeper lying skin layers would not be penetrated by the light to a degree sufficient to excite the PS effectively.

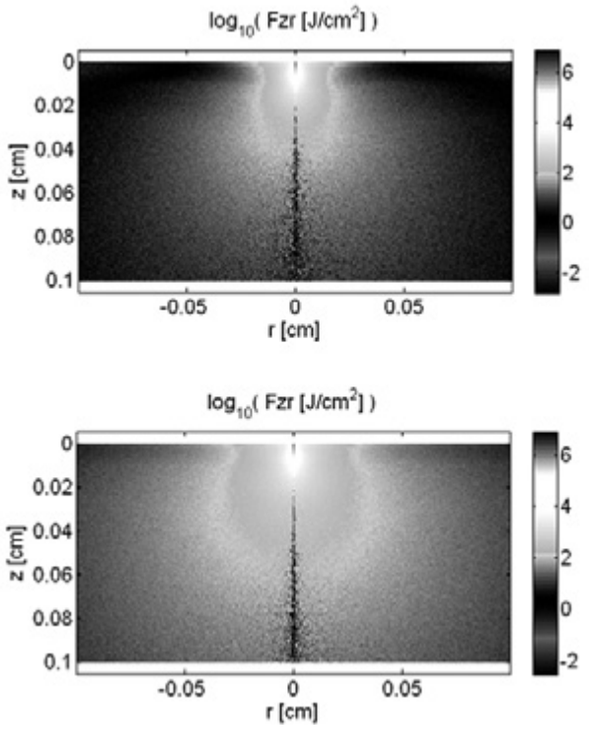


Fig. 5. Monte Carlo simulation of fluence rate distribution (photons λ=650 nm) in seven layer skin model with different content of melanin. Upper: The strongly melanized tissue (30,70,15). Lower: Normal skin (8,15). Parameters of simulation are given in Table.1. The degree of invasion, is taken into account by change of absorption coefficient value of the layer in Table 1. New values of μ_a (in cm^{-1}) are shown in parentheses, first value refers to stratum corneum, second to epidermis, third to papilary dermis, other values of μ_a, μ_s g and n remain unchanged. Fluence rate Frz (r,z- cylindrical coordinates, z-depth, r-lateral position) is given in logarithmic scale. Monte Carlo simulation were performed for infinitely narrow beam (point source), incident light beam perpendicular to tissue surface positioned (0 in z, r scale). Number of photons 10^6, cylindrical symmetry (z, r) used, resolution: step dz=$2*10^{-5}$ (1000 steps), dr $2*10^{-4}$ (500 steps).

The strongly nonhomogeneous distribution of the light intensity changes the excitation conditions of the PS inside the tissue, so the ability of cell killing by PDT inside the cancer volume. Some cell will be destroyed, some not, depending on the light fluence rate values. The influence of the light fluence rate distribution on the cell killing during PDT ($^1O_2$ production) and on kinetics of PDT process (eqs. 13-18) was examined using photochemical parameters of the PS tritolylporphyrin dimer (TD).

TD was chosen as a test compound because it was used earlier as a potentially effective PDT drug and the parameters essential to conduct the kinetic simulations are known for this compound [9]. TD has a relatively large extinction coefficient at 650 nm (ε =$7.55x10^3M^{-1}cm^{-1}$), compared with standard compound usually applied in PDT such as Photofrin (ε =$1.2x10^3M^{-1}cm^{-1}$ at

630 nm [36]). Calculation results for TD are shown in Fig. 6. Simulations were being conducted for a few values of excitation rate constants k_0 corresponding to doses changing from 10 Jcm^{-2} to 80 Jcm^{-2} (at irradiation 100 $mWcm^{-2}$). Inside the melanoma the light fluence rate is attenuated by the presence of additional amount of melanin very strongly (Fig. 5A). In the analyzed model this situation is represented by different k_0 values reflecting changes of fluence rate inside the tissue. Influence of k_0 value on kinetics of $^1O_2$ creation and on killing cells is shown in Fig. 6C and 6D respectively. As described in Fig. 6D, to achieve the threshold level of generated $^1O_2$ we need the relatively high value of k_0. For instance, k_0=66.7s^{-1} (Fig. 6D) is too low and not all the cells in irradiated volume are killed by generated $^1O_2$. The irradiation process should be repeated for such unkilled cells, but the success depends also on the oxygen availability and effectiveness of its diffusion to the place of photosensitizing reaction because the exhaustion of $^3O_2$ finishes the therapy.

Comparison of the results pictured in Fig. 6A and Fig. 6B shows that the relative concentrations of PS and $^3O_2$ plays the key role in PDT efficiency. In a case of high k_0 values, the $^3O_2$ consumption and photobleaching process run practically simultaneously, but in the case of low k_0 value $^3O_2$ depletes as a first (Fig. 6B).

The above described time behavior of PS, $^3O_2$, $^1O_2$ and CT_{ox} shows that from the PDT efficiency point of view the situation of irradiated melanoma cells in the various places of cancer volume could be treated in the simple way considering the k_0 values dependence on melanin content, and therefore the determination of its quantity is a crucial issue.

Since, as mentioned in the Introduction, the evaluation of melanoma is of mainly visual character, the solution of the problem demands the quantitative determination of amount of melanin in tissue.

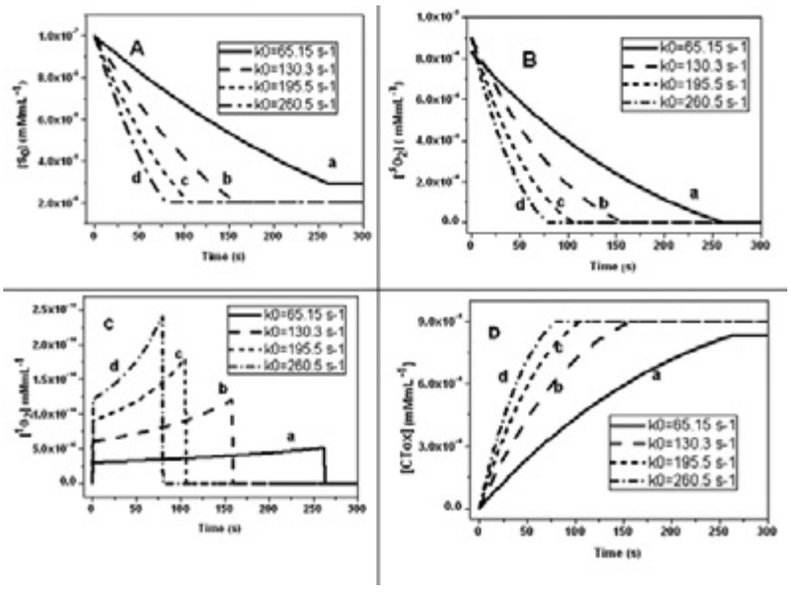


Fig. 6. Simulation of fluence rate influence on time course of some species engaged in PDT. A. $[S_0]$ – concentration of PS in ground state. B. $[^3O_2]$ – oxygen concentration, C. $[^10_2]$ –singlet oxygen concentration, D. $[CT_{ox}]$ – oxidized cellar target concentration. Fluence rate values: k_0=66.16,130.3, 196.6 and 260.6 s^{-1} corresponding to fluence rate φ= 50, 100, 150 and 200 $mWcm^{-2}$ respectively. Calculations were performed using COPASI system [22]. Parameters of simulation: Parameters of simulation: k_f=2x10$^7s^{-1}$, k_p=1x10$^7s^{-1}$, k_L=1x10$^6s^{-1}$, k_{ISC}=8x10$^7s^{-1}$, k_{Pb}=1.2x10$^{12}mMmL^{-1}$, k_T=1x10$^8mMmL^{-1}$, k_{ox}=1.2x10$^{12}mMmL^{-1}$ [20], [CT]=10^{-6}M. Photosensitizer: tritolyloporphyrin dimer (TD), λ_{ex}=649 nm, ε_{649}=7.55x10$^3M^{-1}cm^{-1}$, [S_0 (t=0)]= 10^{-7}M, [$^3O_2$ (t=0)]= 9x10$^{-4}mMmL^{-1}$ [9].

In earlier works a quantitative relationship between the intensity of the reflectance spectrum of melanin and its content in a solid solution was experimentally established [37]. In the considered model a similar relationship was established. The performed Monte Carlo simulations of diffusion reflectance have shown that the intensity of reflectance depends exponentially on the absorption coefficient of melanin (at λ=650 nm), that is on melanin content, and in consequence on the degree of the development of melanoma (Fig. 7). Performing measurements of the future patient skin reflectance spectra (normal and melanized), it would be possible to estimate the melanization degree of the lesion. The knowledge of the melanin content would facilitate the decision on the selection of irradiation parameters (power of light sources and the time of irradiation). which would give, at least theoretically, the certainty of killing all cancer cells.

A separate issue mentioned above is maintaining the oxygen and PS on the level guaranteeing the quantity of generated $^1O_2$ appropriate to kill cells. In the model, concentration of PS and $^3O_2$ was assumed to be constant.

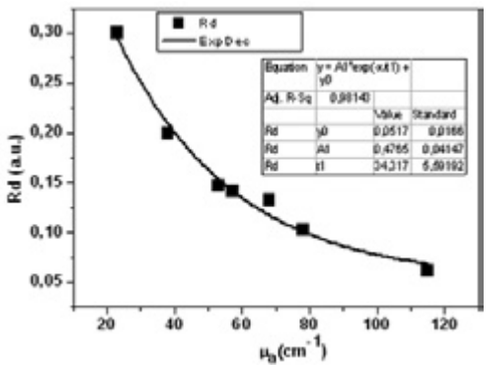


Fig. 7. Monte Carlo simulation of skin diffuse reflectance R_d as function of absorption coefficient μ_a in seven-layer skin model. The values of μ_a depend on melanin content in skin layers. Parameters of simulation are the same as in Fig. 4. Strong dependence of reflectance on absorption coefficient provides the possibility of diagnosis of the degree of melanoma development on basis of reflectance measurement.

Conclusions

The approach discussed in the work has a general character, as it takes into account two main aspects of photodynamic therapy: photon density which constitutes the intensity of primary stimulus inside the tissue and the local photochemical processes generated by absorbed light. The first one depends on melanin distribution inside tissue. Results of simulations for the model discussed in the work provide some useful criteria for improving the PDT therapy. The model not only suggests the potentially optimal anti-melanoma PS compounds but also allows us to calculate the distribution of light intensity and estimate k_0 values inside the tissue when the optical parameters of tissue are known for given absorption properties of the PS used. In these conditions the model could predict approximately the PDT efficiency. Nevertheless, the PDT model considered should be seen as a preliminary attempt towards the study of clinically useful PDT dosimetry.

To improve the predictive abilities of the model, some additional properties should be taken into consideration: the spatial distribution of PS concentration inside the tissue and the problem of oxygen concentration and its transport. The PS and $^3O_2$ concentrations may be heterogeneous inside the tissue and their stationary concentrations are determined through their photochemical reactions and transport. A conclusion to be drawn from this observation is that general simulation of PDT for a specific site in the tissue should be treated as a problem of dynamic optimization, because the rates of photochemical reactions at various sites inside tissue simultaneously depend on volume distributions of PS and $^3O_2$. The kinetic analysis of photooxidation processes leading to cell killing should be also included. In the completed model, the influence of $^3O_2$ transport on the kinetics of $^1O_2$ transport, that was omitted in the previously described model, should be considered in the first place. This requires new experimental data.

References

1. Bajcar S., Grzegorczyk L.: Atlas of skin pigmented lesions, Wydawnictwo Uniwersytetu Jagiellonskiego: Kraków 2000 (in polish).
2. Morton C.A., MacKie R.M.: Clinical accuracy of the diagnosis of cutaneous malignant melanoma. *Br. J. Dermatol.* 1998, 138: 283-287.
3. Abbasi N.R., Shaw H.M., Rigel D.S., Friedman R.J., McCarthy W.H., O sman I., Kopf A.W., Polsky D.: Early diagnosis of cutaneous melanoma: Revisiting the ABCD criteria. *JAMA* 2004, 292: 2771-2776.
4. Morton C.A., MacKie R.M.: Clinical accuracy of the diagnosis of cutaneous malignant melanoma. *Br. J. Dermatol.* 1998, 138: 283-287.
5. Ochsner M.: Photophysical and photobiological processes in the photodynamic therapy. *J. Photochem. Photobiol. B.: Biol.* 1997, 39: 1-18.
6. Young A.R.: Chromophores in human skin. *Phys. Med. Biol.* 1997, 42: 789-802.
7. Urbańska K., Romanowska-Dixon B., Matuszak Z., Oszajca J., Nowak-Śliwińska P., Stochel G.: Indocyanine green as a prospective sensitizer for photodynamic therapy of melanomas. *Acta Biochim. Polon.* 2002, 49: 387-391.
8. Kukielczak B., Cieszka K.A., Matuszak Z.T., Subczynski W.K.: Experimental Photodynamic Therapy (PDT) of Bomirski Hamster Melanoma using Merocyjanine 540 and Visible Light Part I. Pigmented (ma) line. *Curr. Topics in Biophys.* 1995, 19: 66-70.
9. Drzewiecka A., Urbanska K., Matuszak Z., Pineiro M., Arnaut L.G., Habdas J., Ratuszna A., Stochel G.: Tritolylporphyrin dimer as a new potent hydrophobic sensitizer for photodynamic therapy of melanoma. *Acta Biochim. Polon.* 2001, 48: 277-282.
10. Liu L.H.S., Ni C.: Hematoporphyrin phototherapy for experimental intraocular malignant melanoma. *Arch. Ophthalmol.* 1983, 101: 901-903.
11. Gonzalez V.H., Hu L.K., Panagiotis G., Theodossiadis P.G., Flotte T.J., Evangelos S., Gragoudas E.S., Young L.H.Y.: Photodynamic Therapy of Pigmented Choroidal Melanomas. *Invest. Ophthalmol & Visual Sci.* 1995, 36: 871-878.
12. Jacques S.L.: Light distributions from point, line and plane sources for photochemical reactions and fluorescence in turbid biological tissues. *Photochem. Photobiol.* 1998, 67: 23-32.
13. Axelsson J., Swartling J., Andersson-Engels S.: In vivo photosensitizer tomography inside the human prostate. *Optics Letters* 2009, 34: 232-234.
14. Foster T., Murant R., Bryant R., Knox R., Gibson S., Hilf R.: Oxygen consumption and diffusion effects in photodynamic therapy, *Radiation Research* 1991, 126: 29–303.
15. Chandrasekhar S.: Radiative Transfe. *Dover Publications*, June 1960.
16. Niedre M.J., Secord A.J., Patterson M.S., Wilson B.C.: In vitro tests of the validity of singlet oxygen luminescence measurements as a dose metric in photodynamic therap. *Cancer Res.* 2003, 63: 7986-7994.
17. Mishchenko M.I., Travis L.D.d Lacis A.A.: Scattering, absorption, and emission of light by small particles. Cambridge University Press 2002.
18. Sobol I.M.: Monte Carlo Metho. *Nauka*, Moscow 1985.
19. Lakowicz J.R.: Principles of Fluorescence Spectroscopy, Sec. Ed., Kluwer Acad./Plenum Publ. New York 1999.
20. H, X.H., Feng Y., Lu J.Q., Allison R.R., Cuenca R.E., Downie G.H., Sibata C.H., Modeling of a Type II Photofrin-mediated Photodynamic Therapy Process in a Heterogeneous Tissue Phanto. *Photochem. Photobiol.* 2005, 81: 1460--1468.
21. Hoops S., Sahle S., Gauges R., Lee C., Pahle J., Simus N., Singhal M., Xu L., Mendes P. Kummer U.: COPASI – a COmplex PAthway SImulato. *Bioinformatic*. 2006, 22: 3067-3074.
22. Tuchin V.V,: Light scattering study of tissu. *Physics Uspekhi* 1997, 40: 49-515.
23. Cheong W.-F., Prahl S.A., Welch A.J.: A review of the optical properties of biological tissue. *IEEE J. Quant. Electron.* 1990, 26: 2166-2185.
24. Jacques S.L.: Skin optic., Oregon Medical Laser Center News, http://omlc.ogi.edu/news/jan98/skinoptics.htm., Januar 1998.

25. Gemert Van M.J.C., Jacques S.L., Sterenborg H.J.C.M., Star W.M.: Skin Optics. *IEEE Trans. Biomed. Engn.* 1989, 36: 1146-1954.
26. Anderson R., Parrish J.: The optics of human skin. *J. Invest. Dermatol.* 1981, 77: 1-19.
27. Thody A., Higgins E., Wakamatsu K., Ito S., Burchill S., Marks J.: Pheomelanin as well as eumelanin is present in human dermi. *J. Invest. Dermatol.* 1991, 97: 34-344.
28. Megielinsky I.V., Matcher S.J.: Modelling the sampling volume for skin blood oxygenation. *Med. Biol. Engin. Comp.* 2001, 39: 4-49.
29. Meglinsky I.V., Matcher S.J.: Computer simulation of the skin reflectance spectr. *Comp. Meth. Prog. Biomed.* 2003, 70: 17-186.
30. Patwardhan S.V., Dhavan A.P., Relue P.A.: Monte Carlo Simulation of Light-Tissue Interaction: Three-Dimensional Simulation for Trans-Illumination-Based Imaging of Skin Lesions, *IEEE Trans. Biomed. Engn.* 2005, 52: 1227-1236.
31. Meglinsky I.V., Matcher S.J.: Quantitative assessment of skin layers absorption and skin reflectance spectra simulation in the visible and near-infrared regions. *Physiol. Mesur.* 2002, 23: 741-753.
32. Wang L.H, Jacques S.L., Zheng L.Q.: MCML – Monte Carlo modeling of light transportation in multi-layered tissue. *Comp. Methods. Prog. Biomed*. 1995, 47: 131-146.
33. Wang L.H., Jacques S.L., Zheng L.Q.: CONV- convolution for responses to a finite diameter photon beam incident on multilayered tissues. *Comp. Methods. Prog. Biomed.* 1997, 54: 141-150.
34. Zilberstein J., Schreiber S., Bloemers M.C.W.M., Bendel P., Neeman M., Schechtman E., Kohen F., Scherz A., Salomon Y.: Antivascular Treatment of Solid Melanoma Tumors with Bacteriochlorophyl–serine-based Photodynamic Therap. *Photochem. Photobiol.* 2001, 73: 25-266.
35. Gonzalez V.H., Hu L.K., Panagiotis G. Theodossiadis P.G, Flotte T.J., Evangelos S., Gragoudas E.S.. Young L.H.Y.: Photodynamic Therapy of Pigmented Choroidal Melanomas, Invest. Ophthalmol. & Visual Sci., 36: 871-878, 1995.
36. Bergh van den H., Ballini J.-P.: Principles of PDT. In: PDT of Ocular Diseases, Gragoudas E.S., Miller J.W., Zografos L. (eds), Lippincott Williams & Wilkin. 2004.
37. Matuszak Z., Wasilewska-Radwanska M., Buchala J.: The optical properties of some model systems containing melanin, 80 lat Akademii Górniczo-Hutniczej. Proceedings of the Conference. *Physics and its Applications* 1999, pp. 61-66.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 95-100

FINGERPRINT CLASSIFICATION USING COMPUTATIONAL INTELLIGENCE ALGORITHMS IN MEDICAL DIAGNOSTICS

Wiesław Wajs[1], Hubert Wojtowicz[2], Jolanta Wojtowicz[3]

[1] Akademia Górniczo Hutnicza, Wydział EAiE; [2]Katedra Automatyki, Uniwersytet Rzeszowski, Wydział Matematyczno – Przyrodniczy, [3]Katedra Informatyki
hubert.wojtowicz@gmail.com

Abstract: Fingerprint classification is used by anthropologist in detection of genetic disorders in infants. This paper describes application of image processing and pattern recognition methods in classification of fingerprints. Fingerprint classifiers, which are part of an automatic system for rapid screen diagnosing of trisomy 21 (Down Syndrome) in infants, are created and discussed. The system is a tool supporting medical decision by automatic processing of dermatoglyphic prints and detecting features indicating presence of genetic disorder. Images of dermatoglyphic prints are pre-processed before the classification stage to extract features analyzed by the Support Vector Machines algorithm. Application of an algorithm based on multi-scale pyramid decomposition of the image is proposed for the ridge orientation calculation. RBF and triangular kernel types are used in the training of SVM multi-class systems generated with the one–vs–one scheme. The experiments conducted on the database of the Collegium Medicum Jagiellonian University in Cracow show the effectiveness of the proposed approach in classification of infants' fingerprints.
Keywords: image enhancement, fingerprint classification, support vector machines, medical decision support system

Introduction

Early detection of genetic disorders in infants, allowing for making therapeutic decision and starting a treatment, is largely dependent on the ability of carrying out fast and reliable observations and obtaining results of these observations. One of the methods of detecting genetic disorders is dermatoglyphic analysis carried out by an anthropologist. Dermal ridge patterns on fingers, palms and soles used in this method become visible at about three months and are completed by the sixth month of prenatal development. Factors disturbing normal development of fetus may also influence formation of dermal ridges structures. Down syndrome (trisomy 21), one of the most common chromosome disorders and Turner syndrome can be detected using dermatoglyphic patterns analysis [9, 13].

Dermatoglyphic patterns of infants with genetic disorders differ from the normal patterns found in healthy population. To determine the presence of genetic disorder simultaneous analysis of dermatoglyphic prints of fingers, palms and soles is required. Presence of a single pattern typical for the particular genetic syndrome in any of the considered areas is not indicative of Down or Turner syndrome. Many of these patterns can be found in the healthy infants. However, when several, or all, of the patterns characteristic for a genetic disorder are present together, they are indicative of its presence.

For the detection of Down syndrome a diagnostic index was developed called dermatoglyphic nomogram [10]. Diagnostic index used in screening tests for the Turner syndrome presence was also developed [8]. Both of these indexes rely on proper recognition of dermatoglyphic patterns by the anthropologist.

The aim of the work

The aim of the conducted research is to create, basing on gathered data and domain knowledge described in medical literature, an automatic system supporting diagnosis process and detecting infants' genetic disorders, in the following ways:

1. System recognizes characteristic combinations of particular patterns of soles, palms and fingers and on the basis of this recognition, infers about the occurrence of genetic disorders. It is expected, that the application of this system is going to improve treatment effectiveness, i.e. number of complications caused by the treatment in the later years of infants life is going to be lower.

2. System supports doctor's work by the analysis of large amount of patients' data and decreases possibility of making a mistake during strenuous biometric analyses such as counting number of ridges, determining ridges width or calculating ATD angle.
3. System diagnoses presence of the genetic syndromes using image processing methods, pattern recognition and computational intelligence algorithms.

Classification method of fingerprint patterns

Fingerprint classification is one of the tasks of dermatoglyphic analysis. Many classification methods were developed and described in the literature. Classification method used in dermatoglyphic analysis is called Henry method. It classifies fingerprints into five distinct classes called: left loop (LL), right loop (RL), whorl (W), arch (A) or plain arch (PA) and tented arch (TA).

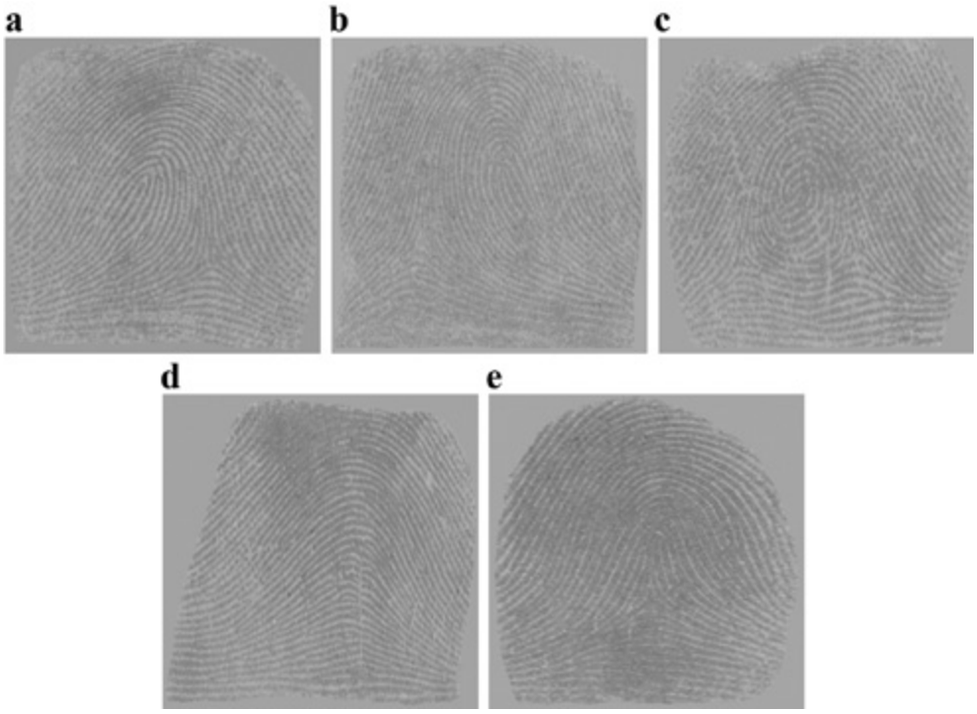


Fig. 1. Example fingerprints: (a) left loop; (b) right loop; (c) whorl; (d) plain arch; (e) tented arch.

Classification scheme based on Henry method is a difficult pattern recognition problem due to the existence of small interclass variability of patterns belonging to the different classes and large intraclass variability of patterns belonging to the same class. Upper row in the figure 2 shows three fingerprint impressions of different topology all belonging to the whorl class, lower row in the figure 2 shows from the left impressions of the fingerprints belonging to the class plain arch, tented arch and right loop.

In the paper we present a classification scheme based on the extraction of fingerprint ridge orientation maps from the enhanced images. Vectors constructed from the directional images are used for the training of the SVM multi-class algorithms. For the induction process of the SVMs RBF and triangular type kernels are used.

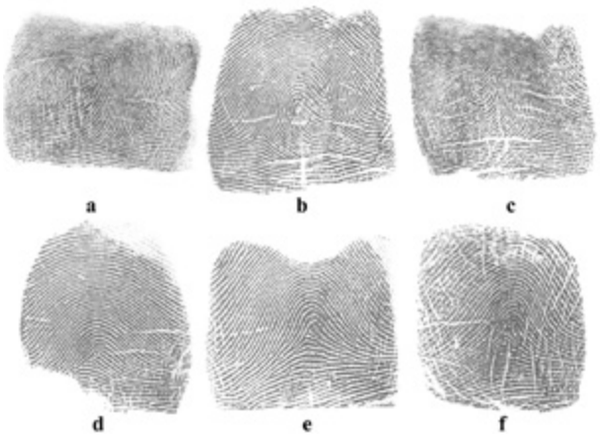


Fig. 2. The upper row contains patterns belonging to the same class but of different topology (large intraclass variability), the lower row contains patterns belonging to the different classes but of similar topology (small interclass variability).

Feature extraction

Accurate extraction of features and classification of fingerprints depends on the quality of the images containing the impressions [6]. Quality of the impressions is not defined by any objective measure. It corresponds to the clarity of the dermal ridges structure. Impressions of good quality are characterized by high contrast and clearly discernible structure of ridges and valleys, poor quality impression are of low contrast and ill-discernible structure of ridges and valleys.

Factors that can adversely affect the quality of impressions are:

– Presence of furrows interrupting ridge continuity. Although high number of furrows on the fingerprint makes the recognition task harder in case of genetic disorders it may be an indication of Down's syndrome.
– Dry fingers yield impressions of fragmentary ridge structure and low contrast.
– Sweat on fingerprints leads to smudges and false connections of parallel ridges.

Fingerprint impressions of low quality complicate the learning process of computational intelligence algorithms, and negatively influence ability to accurately recognize the patterns. Classification accuracy can be improved by image pre-processing of the analyzed images which enhances their quality.

Image pre-processing

Image pre-processing of fingerprint impressions consists of several stages. First stage is an image segmentation, which separates background from the parts of the image where ridges are present.

Segmentation algorithm calculates histogram of the entire image and on the basis of its value a threshold is selected. Areas of the image for which local histogram values are lower than the threshold are treated as a background [7]. After the region mask representing foreground is determined background area is removed from the image. Coordinates of the boundary points of the region mask are found. The image is truncated according to the boundary points coordinates and than using bicubic interpolation resized to the frame of 512 x 512 pixels. In the next stage image contrast enhancement is performed using CLAHE (*Contrast Limited Adaptive Histogram Equalization*) algorithm. CLAHE divides the image into tiles. Each tile contrast is enhanced so that the resulting contrast histogram approximately corresponds to the shape of the statistical distribution specified as an input parameter for the CLAHE algorithm. Adjoining tiles are then combined using bilinear interpolation which smoothes inaccuracies on the edges of the tiles. Contrast in the areas of homogeneous texture can be suppressed in order to prevent amplification of noise always present in some form in the image of fingerprint impression created using offline ink method. Last stage of the pre-processing is an image quality enhancement using a contextual image filtration STFT (*Short Time Fourier Transform*) algorithm [3] which generates information about ridges flow directions, frequency of ridges and local image quality estimation.

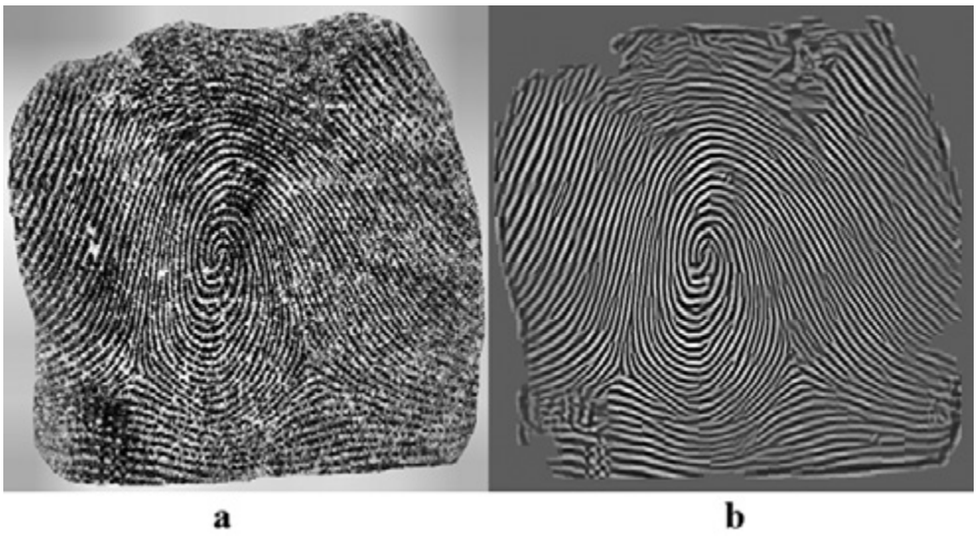


Fig. 3. Image of the fingerprint impression belonging to the whorl class pre-processed using contrast enhancement algorithm CLAHE (a), and then by filtration algorithm STFT (b).

Ridge orientation

Ridge orientation maps are computed using algorithm based on Principal Component Analysis and multi-scale pyramid decomposition of the images [4]. Principal Component Analysis is applied to fingerprint image to calculate maximal likehood estimations of the local ridge orientations. PCA based estimation method uses Singular Value Decomposition of the ensemble of

gradient vectors in a local neighbourhood to find the dominant orientation of the ridges.

Algorithm calculating ridge orientation maps uses also a multi-scale pyramid decomposition of the image, which helps improving accuracy of the estimation. Multi-scale approach consists of following steps:

1. For the computed gradient field for the particular image a Gauss gradient pyramid is created.
2. For every layer of the gradient pyramid, a gradient field is divided into local tiles (overlapping or non-overlapping) and for each tile a local direction is determined using PCA.
3. Estimations are propagated from the lowest resolution level, to the finer levels, up to the highest resolution level

In step 3, the optimal propagation weight is determined by the Kalman gain matrix:

$$K[n] = M[n \mid \gamma n]\,(C[n] + M[n \mid \gamma n])^{-1} \quad (1)$$

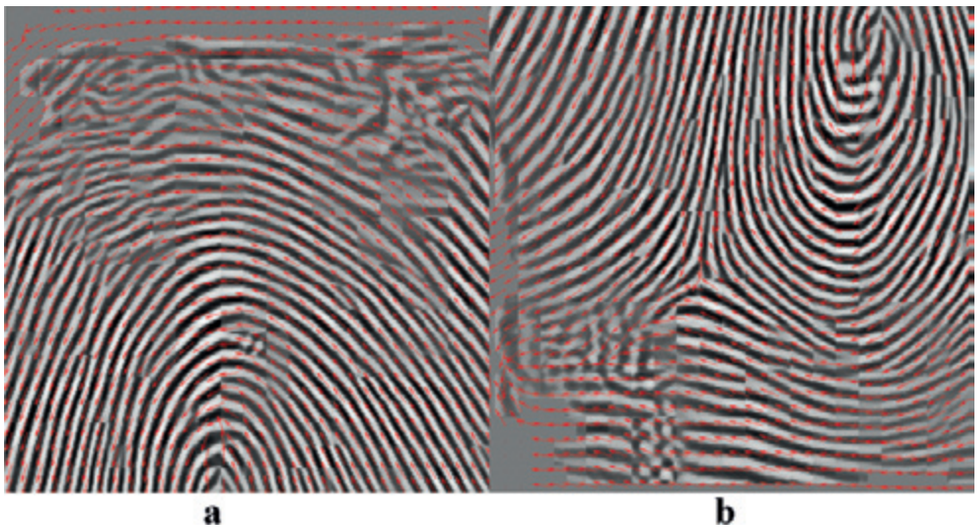


Fig. 4. Local values of ridge orientations in the noised areas of fingerprint impression:
(a) top area of the impression above the upper core of the whorl,
(b) left bottom area of impression containing left triradius and the lower core of the whorl.

Value of the propagation weight in this equation is expressed in terms of covariance matrix of the estimation error $M[n|\gamma n]$ and the covariance matrix of the noise $C[n]$.

Multi-scale method provides robustness to noise present in the image. Examples of the ridge local orientations estimation in the distorted areas of the fingerprint impression from fig. 3 are shown in the fig. 4a and 4b.

Fingerprint classification with Support Vector Machines

In this section SVM algorithm is presented and results of its application to the fingerprint classification task are discussed.

Data

Data set consists of 600 fingerprint 8-bit grey scale images. Size of the images is 512 x 512 pixels. Classes left loop, right loop, whorl arch and tented arch are uniformly distributed in the data set and consist of 120 images each.

A number of images containing partial or distorted information about the class of the pattern for which two or more image copies of the same impression were available were registered using non rigid registration algorithm and then mosaicked.

Classification algorithm

Fingerprint classification was accomplished using a Support Vector Machines algorithm [12]. For multi-class problem an ensemble of SVM classifiers is created trained with 'one-vs-one' voting method. Classifiers are using RBF type kernel functions and triangular kernel functions [5]. Triangular kernel function is formally expressed:

$$k_T(x,x') = -\,\|x - x'\| \quad (2)$$

It has been proven this function is conditionally positive definite [1]. Application of this type of kernel function ensures that the optimized problem is convex and obtained solution is unique [2]. Kernel function is positively definite, if for any x_1, …, x_n and for any c_1, …, c_n such that $\sum_i c_i = 0$, it is satisfied that

$\sum_{i,j} c_i\ c_j\ k_T\ (x_i, x_j) \geq 0$. Due to the equilibrium constraint $\sum_i \alpha_i y_i = 0$, this ensures that k_T can be used as a SVM kernel [11].

Classification results

Training dataset consists of 300 ridge orientation maps calculated from the fingerprint images. It contains 60 maps for each of five the classes. Dataset used for testing of the SVM is comprised of 300 ridge orientation maps. There are 60 maps of LL, RL, W, A and TA classes in the testing set.

Images were in the first stage pre-processed using CLAHE algorithm and then filtered using STFT. Ridge orientation maps were computed from the filter enhanced images using PCA and multi-scale pyramid decomposition algorithm [4].

Cross-validation and grid search methods were used to obtain kernel parameters for the training of the SVM algorithms. Test results of the SVM algorithm with RBF and triangular kernel functions are presented respectively in the fig. 5a and 5b as covariance matrices. Both the SVM trained with RBF kernel and the SVM trained with triangular kernel achieved classification accuracy of 91,3 % on the test data set.

Summary

Experiments described in the paper were performed on the images of varying quality. Most of these images were of low contrast due to the gradual fading of the chemical compound used to create impressions of the infants' fingerprints. Training dataset contained partially blurred fingerprint impressions or impressions

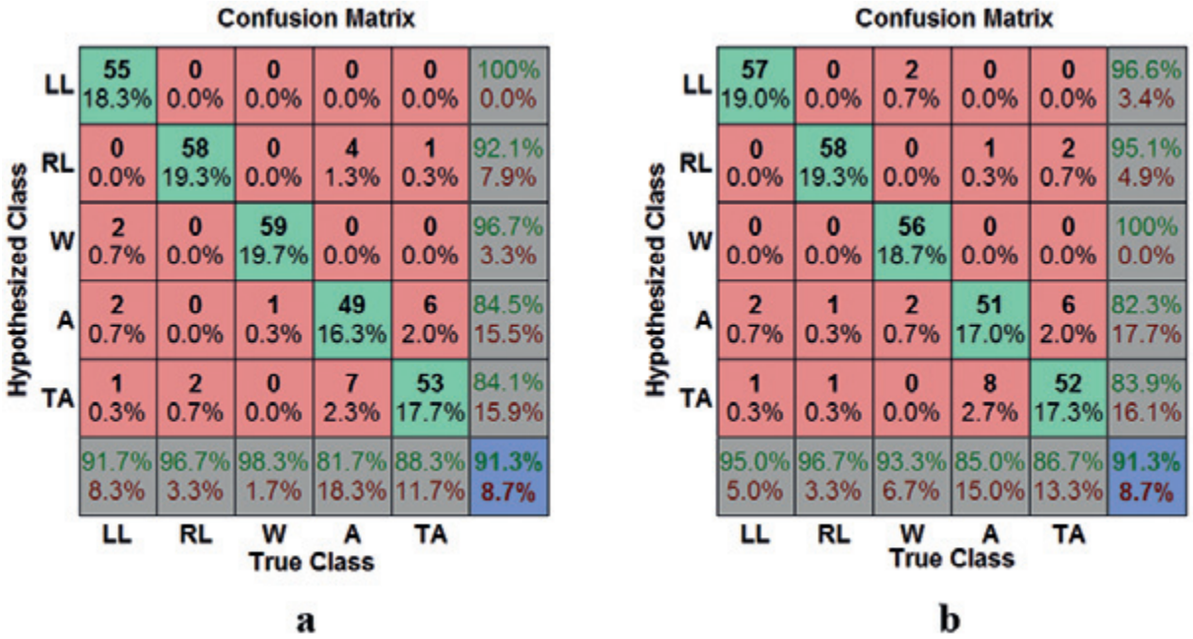


Fig. 5. Test results for the SVM algorithm trained with RBF kernel function (a), test results for the SVM algorithm trained with triangular kernel function (b).

of incomplete patterns. Accurate recognition of ridge directions in the completely blurred areas of the fingerprint is a difficult task. Principal component analysis and multi-scale pyramid decomposition of the image allows for reliable estimation of local ridge directions in the blurred areas, but with the increase of the blurred area size certainty of accurate estimation decreases. Accurate estimation of ridges direction may not be possible if the blurred area contains characteristic points such as cores or triradii. Inclusion of low quality impressions negatively influences selection of the parameters for the training of SVM classifiers and also has a negative impact on the testing accuracy. Achieved classification accuracy is made possible thanks to the pre-processing of the images and application of the RBF and triangular kernel functions in the SVM training scheme

References

1. Berg C.: Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions. Springer–Verlag: Berlin 1984.
2. Boughorbel S., Tarel J. P.: Conditionally positive definite kernels for SVM based image recognition. *IEEE International Conference on Multimedia and Expo* (*ICME)* 2005: 113-116.
3. Chikkerur S., Cartwright A.N., Govindaraju V.: Fingerprint enhancement using STFT analysis. *Pattern Recognition* 2007, vol. 40: 198-211.
4. Feng X. G., Milanfar P.: Multiscale principal components analysis for image local orientation estimation. *The 36th*

Asilomar Conference on Signals, Systems and Computers 2002, vol. 1: 478-482.
5. Fleuret F., Sahbi H.: Scale-invariance of support vector machines based on the triangular kernel. *Proceedings of the Workshop on Statistical and Computational Theories of Vision of the IEEE International Conference on Computer Vision (ICCV/SCTV)* 2003, online.
6. Maltoni D., Maio D., Jain A.K., Prabhakar S.: *Handbook of Fingerprint Recognition*. Springer–Verlag: Berlin 2003.
7. Otsu N.: A threshold selection method from gray-level histogram. *IEEE Transactions on Systems, Man and Cybernetics* 1979, vol. 9, no. 1: 62-66.
8. Preuss M.: A screening test for patients suspected of having Turner syndrome. *Clinical Genetics* 1976, no. 10: 145--155.
9. Reed T.: Dermatoglyphics in Down's syndrome. *Clinical Genetics* 1974, no. 6: 236.
10. Reed T.E., Borgaonkar D.S., Conneally P.M., Yu P., Nance W.E., Christian J.C.: Dermatoglyphic nomogram for the diagnosis of Down's syndrome. *J. Pediat*. 1970, no. 77: 1024--1032.
11. Scholkopf B.: The kernel trick for distances. *Proceedings of Neural Information Processing Systems* (*NIPS*) 2000: 301-307.
12. Steinwart I., Christmann A.: *Support Vector Machines*. Springer Science, 2008.
13. Tornjova-Randelova S.G.: Dermatoglyphic characteristics of patients with Turner's syndrome. *Medicine Anthropologie* 1990, vol. 43, no. 4: 96-100.
14. Zuiderveld K.: Contrast Limited Adaptive Histogram Equalization, Graphics Gems IV. Academic Press, 1994.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 101-110

HEART FAILURE ONTOLOGY

Alan Jovic[1], Dragan Gamberger[2], Goran Krstacic[3]

[1]Faculty of Electrical Engineering and Computing, University of Zagreb,
Unska 3, 10000 Zagreb, Croatia; e-mail: alan.jovic@fer.hr
[2]Rudjer Boskovic Institute, Bijenicka 54, 10000 Zagreb, Croatia; email: dragan.gamberger@irb.hr
[3]Institute for Cardiovascular Diseases and Rehabilitation,
Draskoviceva 13, 10000 Zagreb, Croatia; e-mail: goran.krstacic@zg.t-com.hr
Corresponding author: Dragan Gamberger, Rudjer Boskovic Institute,
Bijenicka 54, 10000 Zagreb, Croatia; Email: dragan.gamberger@irb.hr

Abstract: Ontology represents explicit specification of knowledge in a specific domain of interest in the form of concepts and relations among them. This paper presents a medical ontology describing the domain of heart failure (HF). Construction of ontology for a domain like HF is recognized as an important step in systematization of existing medical knowledge. The main virtue of ontology is that the represented knowledge is both computer and human-readable. The current development of the HF ontology is one of the main results of the EU Heartfaid project. The ontology has been implemented using Ontology Web Language and Protégé editing tool. It consists of roughly 200 classes, 100 relations and 2000 instances. The ontology is a precise, voluminous, portable, and upgradable representation of the HF domain. It is also a useful framework for building knowledge based systems in the HF domain, as well as for unambiguous communication between professionals. In the process of developing the HF ontology there have been significant technical and medical dilemmas. The current result should not be treated as the ultimate solution but as a starting point that will stimulate further research and development activities that can be very relevant for both intelligent computer systems and precise communication of medical knowledge.
Keywords: heart failure, ontology, knowledge representation, decision support system, intelligent data analysis

Introduction

Representation of medical knowledge in a form which enables its use by medical computer-based systems is a long-term goal for clinical researchers and technical people developing such systems. The topic is important due to the complexity of the contemporary medical knowledge and the necessity for unambiguous, consistent, and reliable reasoning using clinical data [1]. Effective knowledge representation is also a hot research topic both in theory and practice of computer science related to artificial intelligence [2].

Ontologies are widely accepted as an appropriate form for the conceptualization of knowledge [3]. They represent a basic step in the knowledge representation process which integrates

a) domain vocabulary (terminology),

b) organization of concepts expressed in the chosen terminology into a hierarchical structure (also known as the taxonomy), and

c) description of relations among concepts and/or classes of concepts.

Ontologies usually do not include procedural knowledge which defines how a specific task can be realized, how some problem can be solved, or what has to be done in a specific situation. Complex manipulations with medical knowledge can be and already have been implemented without ontologies. The examples are diverse expert systems [4-6]. However, construction of ontologies and their usage may have a few decisive advantages:

a) Starting from the available ontology which defines the concepts and relations among them, it is much easier to implement any complex system for data and/or knowledge manipulation. The reason is that using the ontology enables appropriate organization of procedural knowledge, regardless if these procedures and functions are loosely or tightly connected with the ontology, and that can be beneficial for the implementation and maintenance of these systems [1, 7].

b) Ontologies are reusable in various patient data transformations in diverse applications. This facilitates interoperability among the applications, enables easier verification and comparison, and perhaps most significantly, ensures comparability of results coming from applications using the same ontology [8].

c) The development of the ontologies requires that medical concepts are precisely defined. The result may be significant not only for the implementation of technical systems but also as a basis for precise and unambiguous inter-human communication including scientific publications, text books, and guidelines [9-11].

The work in building medical ontologies started with general taxonomies like SNOMED-CT [12] and general ontologies like foundational model of anatomy (FMA) [13]. The main reason for the development of domain-specific ontologies is that general ontologies may not include all domain-related concepts and relations necessary for a particular use case [14]. Another important reason is the level of the granularity of knowledge that is requested by different medical applications. In general, there is no agreement in respect to what type of knowledge a domain-specific medical ontology should include, the optimal organization of the hierarchy of classes, and how relations with general ontologies should be implemented [15].

In spite of significant interest in developing domain-specific ontologies [16], and recognition of the potential benefits of their application [15], there are practically no domain-specific medical ontologies for clinical decision-making that are publicly accepted and widely used [17].

Heart failure (HF) is a one of the leading causes of morbidity and mortality in the world. Despite significant progress in the treatment, HF incidence and prevalence continue to increase, which represents a serious medical problem [18, 19]. There has already been some effort to develop domain-specific medical ontologies. Several approaches have been recently proposed by authors [8, 20, 21].

This paper presents heart failure ontology as one of the main research results of the EU FP6 project Heartfaid [22].

The structure of the work paper is as follows. In section 2 we describe data sources and tools employed in the construction of the ontology. In section 3 we give an in-depth overview of the ontology. Section 4 discusses the significance of the ontology from the standpoint of human communication, patient data analysis, and decision support. We give an overview of the works of other authors in section 5. Section 6 concludes the paper.

Materials and methods

The starting objective for developing the HF ontology was the need for an ontology that will be reusable for different decision support tasks in the heart failure domain. Such an ontology was necessary for the EU Heartfaid project, which required conceptualization and codification of the knowledge from guidelines for handling congestive and acute heart failure patients published by the European Society of Cardiology [19, 23-26]. Other sources of information have been used in its development, including, but not limited to, the Unified Medical Language System (UMLS) [27], the Mayo clinic web site [28] and the Open Clinical web site [29].

The ontology is implemented in the Web Ontology Language (OWL) [30] using the Protégé editing tool [31]. OWL has been developed specifically for web applications (i.e. semantic web) but it also has some other favorable properties: it supports a very expressive description logic which enables specification of complex conditions that may be used for the definition of concepts, it supports both object and datatype properties, and it nicely integrates with Semantic Web Rule Language (SWRL) that can be used for procedural knowledge description in ontologies [32, 33]. There are publicly available open-source reasoners for OWL [34]. The Protégé tool presents a useful and intuitive visual interface for human editing and validation of the ontology and, by its publicly available Java library, it supports and enables development of various software applications that may use the developed ontologies [35]. All these reasons have been decisive for using OWL in this project.

The information contained in the ontology has been added manually. Several currently available automatic tools for ontology construction have been considered in the project, with the most promissing one being Guideline Interchange Format (GLIF) [36]. The automatic tools failed to combine accurately and efficiently the information coming from diverse sources. The content of the ontology has been proof-read by several medical experts collaborating on the project.

The current version of the ontology is in accordance with the last guidelines of European Society of Cardiology for the diagnosis and treatment of acute and chronic heart failure published in the year 2008 [19]. It has been exported into the HTML format and is available as a series of web pages at http://lis.irb.hr/heartfaid/ontology/. Besides for the Heartfaid project, the ontology has been till now used for scientific research purposes by Division of Biomedical Informatics Research at Stanford University School of Medicine [37] and by Heart + Lung Institute at St. Paul's Hospital at University of British Columbia.

Results

Heart failure ontology description

The HF ontology presents the formalized description of concepts for the HF domain. It includes basic HF concepts, properties that characterize patients, all relevant diagnostic examinations and tests, as well as treatment procedures. The ontology also includes other cardiovascular system related concepts as well as concepts related to other organs connected with HF.

The ontology presents a detailed taxonomic overview of the HF domain with around 200 classes describing HF related concepts. Examples are "Cardiac_hypertrophy", "Blood_pressure_signs" or "Heart_murmurs". These concepts are interconnected with super-class and sub-class properties into a hierarchical tree-like structure. At the basic level there are five super-classes: "HF_concept", "Patient_characteristic", "Patient", "Testing", and "Treatment".

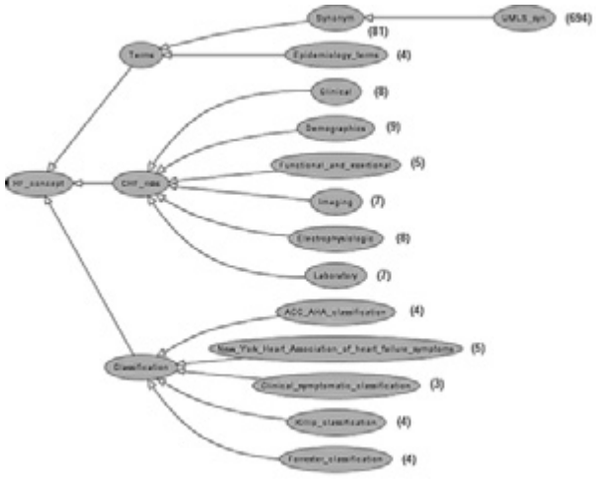
Instances are members of the classes and typically represent list of concrete concepts relevant for the class. For example, the "Cardiac_hypertrophy" class has the following six instances: "Cardiomegaly", "Combined ventricular hypertrophy", "Left_atrial_hypertrophy", "Left_ventricular-hypertrophy", "Right_atrial_hypertrophy", and "Right_ventricular_hypertrophy". In total, the HF ontology includes more than 2000 instances. When possible, instances and classes are connected by their UMLS Concept Unique Identifier (CUI number) and with a list of synonyms. Fig. 1 presents the class "Echocardiography_tests" with seven instances including "Doppler_echocardiography" which has the CUI "C0013520".

Like all ontologies, the HF ontology contains properties that connect instances in different classes. These properties enable representation of relations between two concepts. For example, the relationship between instance "Valvular_heart_disease" from the class "Heart_valve_diseases" and the instance "Dyspnea" from the class of "Signs_and_symptoms" may be represented by the property "CouldBeRelatedTo". Similarly, "Hyperkalemia" from the class "Potassium_disorder" may be linked to "Potassium_sparing_diuretics" or "Spironolactone" by the property "MayBeCausedByMedication". The HF ontology includes definitions of more than 100 different properties. In Fig. 1 we can see that the instance "Doppler_echocardiography" has properties "Definition", "Measures" and "CanDetect". "Definition" is a datatype property which contains a text description of the instance while "Measures" and "CanDetect" are object properties which represent relationships to other instances in the ontology. Specifically, for the instance "Doppler_echocardiography" property "Measures" includes 45 instances starting with "Left_atrial_pressure" while property "CanDetect" includes 9 instances starting with "Right_ventricular_diastolic_dysfunction". Besides that, instance "Doppler_echocardiography" has 13 additional properties that are not presented in Fig. 1.

Heart failure ontology structure

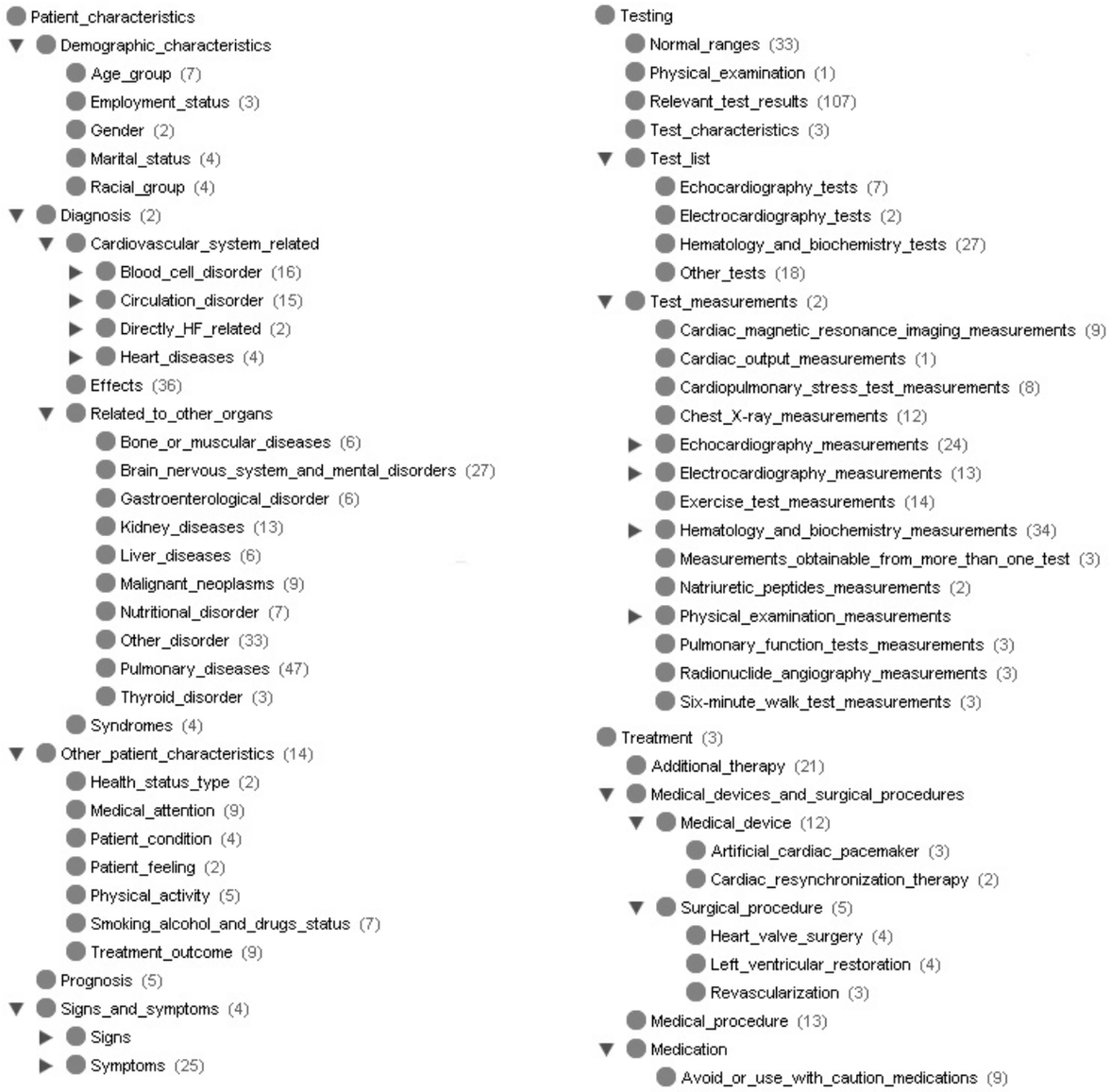
Class "HF_concept" is one of the five main ontology superclasses. It consists of a hierarchy of classes which describe HF terminology, including the risks for congestive heart failure, medical synonyms, and types of classification. One can consider it as a backbone of the whole ontology (Fig. 2). In parentheses, the number of instances in the corresponding class is given.

Class "Patient_characteristics" contains patient's demographical characteristics, possible diagnoses, possible signs and symptoms, prognosis and other characteristics. In fact, this hierarchy defines clinical data in the patient's HF medical record. It is interesting to note that both diagnosis and signs and symptoms have been placed in this class. (Fig. 3). Class hierarchy is shown down to the third-level subclasses due to space limitations.



Class "Testing" represents knowledge regarding tests performed in medical institutions. This includes a list of tests, usual measurements, measurements normal ranges and relevant results. Physical examination has also been placed within this class. Each test relevant to HF has properties that denote the measurements for that test and also which disorders it can detect. Some tests are invasive or used in combination with other tests and this information is also included. The specification of test measurements is as thorough as possible (Fig 4).

Class "Treatment" consists of medical procedures used in the healing process, including medications, devices, invasive and non-invasive procedures, and recommendations regarding HF. Medications are organized into medication groups. Most of the medications relevant for HF symptoms and common comorbidities have been included in the ontology, along with medication dosages and their contraindications (Fig. 5).



The last super-class is “Patient” which is the place reserved for factual knowledge about particular patients. Class “Patient” has no subclasses, so there is no class hierarchy that can be displayed. At the moment, for illustrative purposes, this class contains only three patients, but this is the place where in real applications typically many patient data will be present in the same format. The data may be extracted from medical records or they can be inferred as the result of the reasoning process. Patients have roughly 40 properties in the ontology.

For a more detailed representation of the ontology structure than the ones shown in the Fig. 2-5 we refer the reader to the ontology web site [http://lis.irb.hr/heartfaid/ontology/].

Most significant classes

In this section we present a few most important classes in the HF ontology.

Class "Diagnosis"

Class "Diagnosis" is a subclass of the class "Patient_characteristics". It consists of four subclasses: "Cardiovascular_system_related", "Effects", "Related_to_other_organs", and "Syndromes". "Cardiovascular_system_related" contains also four important subclasses as shown in Fig. 3. These are "Blood_cell_disorder", "Circulation_disorder", "Directly_HF_related" and "Heart_diseases". Each of these classes is further divided into many subclasses and instances.

Class "Heart_diseases" contains the list of all heart-related disorders, excluding HF. For example, instances "Left_atrial_hypertrophy", "Myocardial_fibrosis", "Cardiomegaly", "Aortic_valve_insufficiency", "Sick_sinus_syndrome", "Left_bundle_branch_block", and many others are members of the subclasses of the class "Heart_diseases".

Class "Directly_HF_related" contains very specific diagnoses related directly to HF, such as "Chronic_heart_failure", "Acute_heart_failure", "Left_ventricular_systolic_dysfunction", "Diastolic_heart_failure", and others.

Some of the most interesting classes in the ontology are "Blood_cell_disorder" and "Circulation_disorder", because heart failure is often in direct relation with the dynamics of the blood flow and with its content. Examples of included concepts are: "Hypovolemia", "Sepsis", "Polycythemia", "Thromboembolic_event" and "Hemorrhage". However, there are lot of other important blood and circulation related disorders and not all of them could be included in this ontology.

We added many other disorders that are not cardiovascular because they are related to the functioning of the heart, relevant as possible causes of the HF symptoms, or relevant to the treatment of HF. These are included in the class "Related_to_other_organs". Some of the examples are: "Skeletal_muscle_problems", "Anaemia", "Cerebral_hemorrhage", "Drug_abuse", "Pneumonia", etc.

Some of patient statuses that can not be exactly considered as a diagnosis and some known HF syndromes are given in classes "Syndromes" and "Effects". Examples include: "Lack_of_adequate_sleep", "Meningism", "Overeating" and "Reduced_sudden_death". The class "Diagnosis" thus contains many different possible aspects of the heart failure disorder and even a wider range of diseases. Most of the significant diseases which are considered to be relevant to heart problems in any way are members of this class.

Class "Medication"

"Medication" is a subclass of the root class "Treatment" and it contains HF related medications and medication groups. It also contains some of the other generic medications used in treatment of heart related problems, such as medications for atrial fibrillation or high blood pressure. This class is divided into three classes: "Avoid_or_use_with_caution_medications", "Heart_failure_medication_group", and "Other_medications_ groups".

Classes for the specific medication group include individual generic medications. Medication groups may also have a specific instance, e.g. "ACE_inhibitors" or "Nitrates" or "Angiotensin_II_receptor_blockers". Individual medications and medication groups contain about a dozen important properties such as properties related to dosage: "InitiatingDose", "TargetDose", "MaximumRecommendedDailyDose", "MaintenanceDose", "SideEffect", "Indicated", "Contraindicated", etc. These properties link medications with signs and symptoms, diagnosis and other medications.

Class, "Avoid_or_use_with_caution_medications," contains specific medications that should not be prescribed to the patient if the patient has HF, as recommended by chronic heart failure guidelines, such as "Corticosteroids", "Diltiazem" and "Verapamil". [19, 23-25]. Finally, "Other_medication_groups" includes those groups of medications and individual medications not directly used in the treatment of HF, but rather in the treatment of the most common comorbidities. Currently, there are a total of 37 medication groups and about 100 individual generic medications in the HF ontology.

Class "Testing"

The sub-class "Test_list" contains a thorough list of tests spanned through four classes: "Echocardiography_tests", "Electrocardiography_tests", "Hematology_and_biochemistry_tests" and "Other_tests". There are seven individual tests under "Echocardiography_tests" as shown in Fig. 1. There are two electrocardiography based tests: "Electrocardiogram_at_rest" (12-lead) and "Holter_electrocardiography_24_hour". There are 27 hematology and biochemistry tests, for example: "C-reactive_protein_test", "Complete_blood_count", "Leukocytes", "Lipid_panel", "S-glucose" etc. "Other_tests" include 18 tests, e.g. "Cardiac_MRI", "Cardiovascular_monitoring", "Chest_CT", "6_minute_walk_test" and "Thoracic_radiography". Each test has its measurements specified in a separate class. All of these classes are placed in the class "Test_measurements", under the class "Testing". Test results relevant for inference of some disorders are placed in the separate class "Relevant_test_results", which enumerates 107 instances. For example: "Cardiothoracic_ratio_greater_than_0.5", "BNP_value_higher_than_100_pg_per_ml", "E_A_ratio_less_than_1" etc.

There also exists a separate class "Normal_ranges" used to specify the normal values, most often in mg/dl or mmol/l, but also in other measure units.

Class "UMLS_syn"

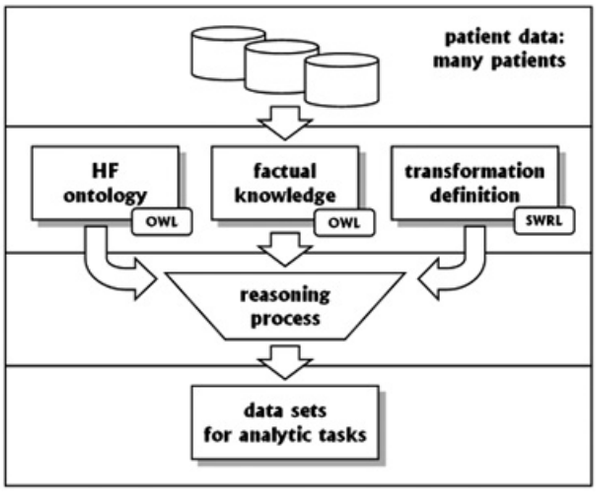
Finally, we consider the class "UMLS_syn", which is a subclass of class "Synonym", subclass of "Terms" and subclass of "HF_concept" (see Fig. 2). Class "UMLS_syn" contains many of the synonyms taken from Unified Medical Language System. In OWL, there can be only one instance with each unique name in the whole ontology despite the fact that there may be many synonyms for any given concept. This is solved by creating instances of significant synonyms in the class "UMLS_syn". Each instance of this class has its name and CUI, which identifies it in UMLS. For example, instance "Dehydration" of the class "Effects" has UMLS synonyms "Exsiccosis" and "Dehydrated" in class "UMLS_syn" and all possess the same CUI: C0011175.

Class "UMLS_syn" differs from a more general class "Synonym" in that an instance located in the class "UMLS_syn" must

be found in the UMLS while instances in superclass "Synonym" need not be present in UMLS. An example is the instance "Acute_heart_failure" in the class "Heart failure". It has synonyms: "Decompensated_heart_failure" and "AHF" in class "Synonym" and "Cardiac_failure_acute" in class "UMLS_syn". "Decompensated_heart_failure" and "AHF" do not exist in UMLS (at present), but are important concepts in guidelines for the acute heart failure and are thus added as synonyms for the acute heart failure (although decompensated heart failure is not strictly a synonym for the acute heart failure, it is its most common case).

Discussion about applications

The primary application of this HF ontology is in human professional communication [9]. For this purpose it is enough to verify that a term exists in the ontology and that the relations as described by the ontology correspond to our understanding of the underlying concept in a concrete situation. If these conditions are met, then we can be fairly certain that other humans will be able to correctly understand our statements. If a term is not in the ontology or its relations to other concepts are different from those that we assume, a good practice is to specify explicitly the term and its relations with those concepts that are present in the HF ontology. If such a concept and its relations are important for many users, it can be added to revised versions of the ontology.



Patient data transformation

Very useful is the application of the ontology for systematic and reproducible patient data analysis tasks. For this purpose, an addition of some procedural knowledge for the transformation or abstraction of patient data is necessary. By connecting these procedural relations with the ontology we ensure that implemented relations can be verified and potentially reused on different datasets. This approach is especially relevant for scientific research tasks in clinical studies because it is possible to ensure transparency and reproducibility of the obtained results [7, 37]. Currently, the most popular is the use of SWRL for coding procedural relation in the form of rules. Publicly available description logic reasoners such as Pellet [34] can be used for the execution of the data transformation and data preparation process. A schematic illustration of this type of application is presented in Fig. 6. The figure demonstrates that the actual data transformation is performed by the reasoning process performed on patient data previously extracted from patient records and presented in the ontology as factual knowledge. The transformation definitions are in the form of SWRL rules which use concepts defined by the HF ontology.

Decision support

The most sophisticated is the application of the HF ontology in intelligent decision support systems (DSS). In this type of application, a complete expert system for patient related warnings, suggestions, and/or decisions may be implemented.

The starting point is descriptive knowledge about the medical domain (HF ontology) into which procedural knowledge (also called actionable knowledge) presented in the form of rules is added [38]. The rules connect some patient properties with relevant conclusions. An example is the rule for systolic HF which states:

Patient has systolic HF if

a) he has performed echocardiography and

b) has either decreased left ventricular contractility or left ventricular ejection fraction below 40% and

c) has some HF signs or symptoms.

SWRL may be used for the presentation of these rules in the same way as described in the data transformation applications. Because of such integration, the procedural knowledge uses the taxonomy of the HF ontology and, what is even more relevant, it may use information from the descriptive knowledge part. An example is that in the case of the rule for systolic HF we do not need to list all HF signs and symptoms in this rule. We only need to test if a patient has any instance from the classes defining HF signs and HF symptoms. Another example is that in the HF ontology there are classifications of medications and their initiating and target doses. This information may be used to implement rules that warn about potentially conflicting situations when there are two instances from the same class for the same patient (for example in the class "Angiotensin_II_receptor_blocker"), to suggest initiating dose when introduction of a medication is suggested, and to warn when current dose is higher than the target dose.

There are two drawbacks that we are aware of when using SWRL as a rule-based system. The first one is that the Pellet reasoner supports only decidable (DL-safe) SWRL rules. This limits the application of rules to only those instances that are contained in the ontology. A solution is a translator between a database and the ontology that transforms database entries into a set of allowed instances. The second problem with SWRL is that negation-as-failure is not supported. This problem can be solved by applying quantifiers on the properties or by introducing instances representing negative results.

In the realized HF expert system for the Heartfaid platform, procedural knowledge is organized in eight groups of rules including: diagnosis, alternative diagnosis, severity assessment, prognosis, medication prescription and medication related warnings, and acute decompensation detection. In total, the expert

system consists of about 200 rules that are in form similar to the presented rule for systolic HF.

The third necessary component of an expert system, besides descriptive and procedural knowledge, is factual knowledge about real patient data. The data are extracted from patient records and inserted into the ontology in the form of instances into the class "Patient". The data may be categorical (like "Patient_performed_echo" or "Patient_has_low_EA_ratio") or they may be numerical (like "Measured_ejection_fraction" equals 30%). In the latter case additional SWRL rules are used to generate categorical instances necessary for decision support tasks (like "Ejection_fraction_above_40", or "Ejection_fraction_30-40"). The conclusions obtained as a result of the reasoning process (like "Patient_has_systolic_HF") are also instances of the class "Patient".

The process and complexity of the transformation of patient data into ontological form of factual knowledge and the process of presentation of results of the reasoning from the ontological form into human interpretable suggestions depend on the particular application. The reasoning part of the expert system may be implemented by general reasoning tools in the same way as for data transformation tasks as shown in Fig. 6. The only differences between the applications for patient data analysis tasks and the decision support applications are that a) the latter have as their input data about one patient only and b) their output is not only the transformed input information but concrete warnings, suggestions, and/or decisions that may be interpreted by humans.

The similarities between these two types of apparently very different applications and the utility of a single HF ontology for both of them, demonstrate the reusability of the implemented ontology. This was a primary motivating factor in choosing to use ontologies as the knowledge-representation method for the EU Heartfaid project. The approach is suggested also for applications in various clinical studies.

In the Heartfaid project we have also made an experiment that is applicable only for decision support applications of ontologies. In this experiment we have introduced actionable subclasses of the class "Patient" that are defined by necessary and sufficient conditions representing rules in the description logic formalism. The approach is illustrated in Fig. 7. The motivation has been to construct a hierarchy of actionable classes in a way similar to the ontology for descriptive knowledge. The goal has been natural ordering and clustering of rules with the aim to enable more effective maintenance of procedural knowledge and more human intuitive coding of rules than in the implementation using SWRL. In this setting the result of reasoning is not introduction of an additional instance like "Patient_has_systolic_HF" but the result of reasoning is placement of the patient, which is already an existing instance and which is the object of reasoning, into the class "Systolic_positive" that is defined with the rule describing systolic HF patients (Fig. 7). The characteristics and problems related to this approach are the topic of further research and are out of scope of this work.

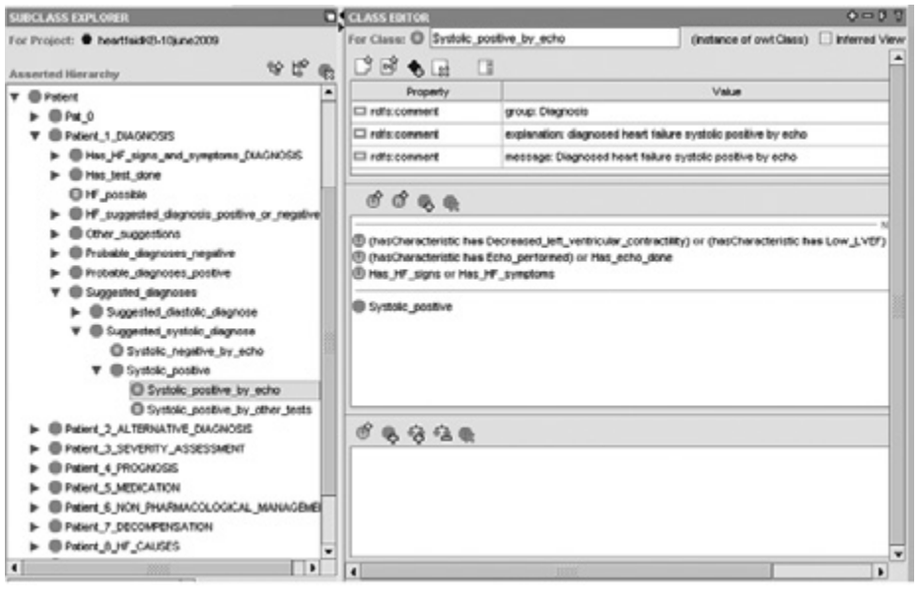
Applications in other domains

The HF ontology presented in this work has already been used for such applications by Stanford University School of Medicine [34]. In this application, the ontology has been used for the research in the field of etiology and risk factors of the nonischemic heart failure. For this application three different knowledge bases have been developed on the top of the HF ontology: a HF pathophysiology knowledge base, a diagnostic criteria knowledge base, and a nonischemic HF etiology knowledge base. These knowledge bases are then applied to patients' clinical data using SWRL for patient phenotype classification, data set generation, and hypothesis validation and discovery in the framework as presented in Fig. 6.

Related work

This section gives an overview of several recently proposed approaches to ontology-based, domain-specific modeling in cardiovascular medicine [8, 20, 21].

The work by Eccher et al. [8] promotes a very flexible architecture for supporting a health care process and its interface with medical knowledge bases for the case of HF. Also interesting is



the approach for building the ontology that starts with archetypes that are developed for the concrete application, which are then integrated into general reference ontology of medicine (DOLCE). The work performed by Esposito [20] deals with ontology-based reasoning applied to patients with congenital heart disease (CHD). By using some specific parts of the SNOMED medical terminology, the author managed to construct a small domain-specific ontology used to detect congenital malformations of the heart and of its blood vessels.

The authors Chiarugi et al. [21] recently proposed a DSS for heart failure patients' management based on the knowledge acquired through continuous collaboration on the Heartfaid project. The smaller ontologies used by the DSS were developed for the purpose of faster computer reasoning as well as for easier maintenance. The authors designed the DSS in such a way that it takes signal and image processing patient data, stores them in a database, transforms and imports the data in the ontologies by using Jena framework, reasons on the data by using Pellet reasoner, and displays various suggestions to medical personnel.

Conclusions

In our work we have been confronted with many dilemmas and the current version of the HF ontology may be interesting also as a prototype for building similar medical ontologies for other domains. With the discussion on potential benefits of using ontologies, we also try to motivate the medical community for stronger participation in building and refinement of medical ontologies. Nobody can better define the ontology than the HF experts, and the heart failure community can significantly profit if both technical systems and humans in their communication start to use this standard.

This HF ontology has been developed mainly by technical people by reading medical literature, primarily HF guidelines published by European Society of Cardiology. In this sense the current version is more an effort to demonstrate how a useful ontology may look like than the final product. In order to stimulate its application and its further development, the complete HF ontology is made public. Constructive criticism may help us to improve the ontology iteratively. How the process of the ontology maintenance should be organized in order to ensure its public relevance and constant improvements at the same time, is still an open issue.

By connecting the procedural tasks with the ontology as part of data preprocessing, we ensure transparency, consistency, and reusability of the procedures, characteristics that are important for scientific research tasks and medical trials. Integration of procedural knowledge (rules) with ontological representation of domain related descriptive knowledge in decision support tasks enables direct use of descriptive knowledge in the decision making process. A nice property of such integration is that the reasoning result may change automatically when the knowledge present in the HF ontology is updated. The approach has been demonstrated as useful within the Heartfaid platform and is suggested also for other similar projects.

Acknowledgements

This work was supported by the EU FP6 project "Heartfaid: A knowledge based platform of services for supporting medical-clinical management of the heart failure within the elderly population" and by the Croatian Ministry of Science, Education and Sport project "Machine Learning Algorithms and Applications". We are grateful to Dr. David Kao from Division of Biomedical Informatics Research at Stanford University School for useful comments.

References

1. de Clercq A., Blom J.A., Korsten H.M., Hasman A.: Approaches for creating computer-interpretable guidelines that facilitate decision support. *Artif. Intell. Med.* 2003, 31(1): 1-27.
2. Negnevitsky M.: Artificial Intelligence: A Guide to Intelligent Systems. 2nd ed. Addison Wesley: Harlow 2005.
3. Pisanelli D.M. (ed.): Ontologies in Medicine. IOS Press: Amsterdam 2004.
4. Barnett G.O., Cimino J.J., Hupp J.A., Hoffer E.P.: DXplain. An evolving diagnostic decision-support system. *JAMA* 1987, 258(1): 67-74.
5. Edwards G., Compton P., Malor R., Srinivasan A., Lazarus L.: PEIRS: a pathologist maintained expert system for the interpretation of chemical pathology reports. *Pathology* 1993, 25: 27-34.
6. Tierney W.M., Overhage J.M., Takesue B.Y., Harris L.E., Murray M.D., Vargo D.L. et al.: Computerizing guidelines to improve care and patient outcomes: the example of heart failure. *J. Am. Med. Inform. Assoc.* 1995, 2(5): 316-322.
7. Charest M., Delisle S.: Ontology-guided intelligent data mining assistance: Combining declarative and procedural knowledge. *Proc. of Artificial Intelligence and Soft Computing* 2006: 9-14.
8. Eccher C., Purin B., Pisanelli D.M., Battaglia M., Apolloni I., Forti S.: Ontologies supporting continuity of care: The case of heart failure. *Comput. Biol. Med.* 2006, 36: 789-801.
9. Serban R., ten Teije A., van Harmelen F., Marcos M., Polo-Conde C.: Extraction and Use of Linguistic Patterns for Modelling Medical Guidelines. *Artif. Intell. Med.* 2007, 39(2): 137-149.
10. Bradley J.H., Paul R., Seeman E.: Analyzing the structure of expert knowledge. *Information and Management* 2006, 43: 77-91.
11. Chi M.T.H., Feltovich P.J., Glaser R.: Categorization and representation of physics problems by experts and novices. *Cogn. Sci.* 1981, 5: 121-152.
12. Spackman K.A.: SNOMED-CT milestones: endorsements are added to already-impressive standards credentials. *Healthc Inform.* 2004, 21: 54-56.
13. Rosse C., Mejino J.L.V.: A reference ontology for bioinformatics: the foundational model of anatomy. *J. Biomed. Inform.* 2003, 36: 478-500.
14. Na J.C., Neoh H.L.: Effectivenes of UMLS semantic network as a seed ontology for building a medical domain ontology. *Aslib. Proceedings* 2008, 60: 32-46.

15. Kumar A., Lina Y., Smith B., Grenon P.: Bridging the gap between medical and bioinformatics: An ontology case study in colon carcinoma. *Comput. Biol. Med.* 2006, 36: 694--711.
16. Smith B. (eds): International Conference on biomedical ontology. 2009 July 24-26, Buffalo, USA 2009.
17. The Open Biomedical Ontologies: OBO Foundry candidate ontologies, http://www.obofoundry.org/.
18. Yan A.T., Yan R.T., Liu P.P.: Pharmacotherapy for Chronic Heart Failure: Evidence from Recent Clinical Trials. *Ann. Intern. Med.* 2005, 142(2): 132-145.
19. Dickstein K., Cohen-Solal A., Gerasimos F., McMurray J.J.V., Ponikowski P., Poole-Wilson P.A. et al.: ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur. Heart J.* 2008, 29(1): 2388-2442.
20. Esposito M.: Congenital Heart Disease: An Ontology--Based Approach for the Examination of the Cardiovascular System. *Lecture Notes in Computer Science* 2008, 5177: 509-516.
21. Chiarugi F., Colantonio S., Emmanouilidou D., Martinelli M., Moroni D., Salvetti O.: Decision support in heart failure through processing of electro- and echocardiograms. *Artif. Intell. Med.* 50(2), 95-104, 2010.
22. Conforti D., Constanzo D., Perticone F., Parati G., Kawecka-Jaszcz K., Marsh A. et al.: HEARTFAID. A knowledge based platform for supporting the clinical management of elderly patients with heart failure. *Journal on Information Technology in Healthcare* 2006, 11: 46-51.
23. Nieminen M.S., Bohm M., Cowie M.R., Drexler H., Filippatos G.S., Jondeau G. et al.: Executive summary of the guidelines on the diagnosis and treatment of acute heart failure: the Task Force on Acute Heart Failure of the European Society of Cardiology. *Eur. Heart J.* 2005, 26: 384-416.
24. Swedberg K., Cleland J., Dargie H., Drexler H., Follath F., Komajda M. et al.: Guidelines for the diagnosis and treatment of chronic heart failure: executive summary (update 2005): The Task Force for the Diagnosis and Treatment of Chronic Heart Failure of the European Society of Cardiology. *Eur. Heart J.* 2005, 26: 1115-1140.
25. Remme W.J., Swedberg K.: Guidelines for the diagnosis and treatment of chronic heart failure. *Eur. Heart J.* 2001, 22: 1527-1560.
26. Cleland J.G., Swedberg K., Follath F., Komajda M., Cohen--Solal A., Aguilar J.C. et al.: The EuroHeart Failure survey programme – a survey on the quality of care among patients with heart failure in Europe. Part 1: patient characteristics and diagnosis. *Eur. Heart J.* 2003, 24: 442-463.
27. Unified Medical Language System, http://www.nlm.nih.gov/research/umls/.
28. Diseases and Conditions – MayoClinic, http://www.mayoclinic.com/health/DiseasesIndex/DiseasesIndex.
29. OpenClinical: knowledge management technologies for healthcare, http://www.openclinical.org/.
30. Web ontology Language OWL / W3C Semantic Web Activity, http://www.w3.org/2004/OWL/.
31. Noy N.F., Sintek M., Decker S., Crubezy M., Fergerson R.W., Musen M.A.: Creating Semantic Web Contents with Protégé-2000. *IEEE Intelligent Systems* 2001, 16(2): 60-71.
32. SWRL: A Semantic Web Rule Language Combining OWL and RuleML, http://www.w3.org/Submission/SWRL/.
33. Shadbolt N., Berners-Lee T., Hall W.: The Semantic Web Revisited. IEEE Intelligent Systems 2006, 21(3): 96-101.
34. Pellet: The open source OWL reasoner, http://clarkparsia.com/pellet/.
35. Protégé-OWL API, http://protege.stanford.edu/plugins/owl/api/.
36. Ohno-Machado L., Gennari J.H., Murphy S.N., Jain N.L. et al.: The GuideLine Interchange Format. *J. Am. Med.* Inform. Assoc. 1998, 5: 357-372.
37. Kao D.P., Pavlovic A., Fowler M.B., Das A.K.: Taking Historical Inventories in Nonischemic Heart Failure – Etiologies and Risk Factors Trial (THIN-HEART) – The Heart Failure Origin Organization Task (HOOT). Proc. of Summit on Translational Bioinformatics, San Francisco, CA 2008, poster.
38. Prcela M., Gamberger D., Jovic A.: Semantic Web Ontology Utilization for Heart Failure Expert System Design. Proc. of MIE2008 Int. Congress of the European Federation for Medical Informatics 2008, pp. 851-856.

Figure captions

Fig. 1. Instances in class “Echocardiography_tests” and some properties of the instance “Doppler_echocardiography”. It can be seen that the instance has CUI and “Definition” as obligatory parts and some other properties like “Measures” and “CanDetect”.

Fig. 2. The “HF_concept” super-class includes following subclasses: “Terms”, “CHF_risks”, and “Classification”.

Fig. 3. “Patient_characteristics” super-class consists of: “Demographic_characteristics”, “Diagnosis”, “Other_patient_characteristics”, and “Signs_and_symptoms”. Each of these classes includes many different concepts.

Fig. 4. The hierarchy of the class “Testing”.

Fig. 5. Super-class “Treatment” consists of three sub-classes: “Additional_therapy”, “Medical_devices_and_surgical_procedures”, and “Medication”. This last one is relevant because of the information about many specific medications.

Fig. 6. Application of the ontology for systematic and reproducible patient data analysis tasks.

Fig. 7. Procedural knowledge integrated into HF ontology as actionable classes for the implementation of the Heartfaid expert system. The central part of the figure presents definition of a rule for the diagnosis of the systolic heart failure by using description logic. The left part of the figure presents eight super-classes in which actionable knowledge is included in a hierarchical order.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY JAGIELLONIAN UNIVERSITY – MEDICAL COLLEGE
Vol. 7, No. 2, 2011, pp. 111-116

THE APPLICATION OF THE JSP METHOD TO THE SYSTEM DESIGNED FOR THE MANAGEMENT OF A CHOSEN MEDICAL UNIVERSITY DEPARTMENT

Piotr Ziniewicz[1], Robert Milewski[1],
Paweł Malinowski[1], Stanisław Zenon Mnich[1]

[1]Department of Statistics and Medical Informatics, Medical University of Bialystok, Białystok, Kilińskiego 1, Poland

Abstract: Dynamic advances in medical computer science and growing demands of computer users lead to the creation of novel computer systems of increasing complexity. Such systems contain a number of co-operating components which are located on many machines and communicate with each other using various methods. The process of designing and creating such complex system is a great software engineering challenge. JSP is one of many methods that can cope with the complexity, and organize programmers team work.
Keywords: software construction, computer systems development, JSP method

Dynamic advances in medical computer science and growing demands of computer users lead to the creation of novel computer systems of increasing complexity. Such systems contain a number of co-operating components which are located on many machines and communicate with one another using various methods. Applications designed to realize single tasks are slowly becoming relics of the past, being replaced by more difficult and complex systems designated to manage in an organized and guided manner the work of an individual. This means that algorithm creation skills have to be complemented with the knowledge related to systematic design of complex computer systems, especially when they are created by an organized team of programmers. Then, not only the requirements of project realization, but also proper communication within the team should be defined. An additional problem involves changes in the system specification that may alter its structure. Taking the above into consideration, it can be concluded that the lack of experience with modern methods of software design is one of the most frequent reasons for the failure of the entire process [4].

The basic problem that appears during software construction is its complexity [1]. When the system becomes too complex, the designer ceases to "reign" over its every aspect. The general understanding of such a complex system in its all aspects exceeds human possibilities. The division (structuralization) of the system according to the defined criteria seems to be the solution to the problem. However, the problem described by the system has to meet the basic assumption underlying its decomposition, otherwise structuralization of the system will not be possible.

The decomposition means that every element can be described by the sequence of more detailed elements (sub-systems) which also can be split to even simpler elements. This process should be continued until the obtained elements are trivial to implement. As a result of this process, the hierarchical structure is achieved. It is assumed that in such sub-systems, inner-group communication is more dynamic than the outer one. Expanding functionality is the natural tendency of computer systems. Decomposition should be conducted in the way that enables easy implementation of a wide range of changes and other modules to the already existing code.

While constructing the computer system it is assumed that the problem to be solved with its use can be decomposed. The software construction is divided into 5 stages:

- Analysis
- Design
- Implementation
- Testing
- Maintenance

The analysis stage includes formalization of the set of system requirements to meet, construction of the functional model of this system, building a model of system interactions with the external world (users, machines) and creation of the information flow

control model between the system and its modules. The design stage translates the logical description defined in the previous stage into the description of the physical structure of the system. Based on the physical structure, elements of hardware can be defined: computers, servers, network media; software: applications, objects, functions; and their allocation to equipment components. The implementation stage means coding the algorithms and data, using the specific programming language to create a system structure that matches the earlier physical description. In the testing stage software components have to be physically allocated to the hardware units described by the system structure. At this stage, proper functioning of the software is tested in the user's environment. Any bugs that appear are fixed. The ultimate user can have access to the system during this phase. Finally, the maintenance stage begins from the moment of the system delivery to the user and involves: continuous elimination of the bugs that come out during exploitation, expanding the system functionality according to the user's suggestions and improving the overall performance. Each of these stages should be realized by an individual team. The stages may intervene or can be implemented consecutively.

The analysis and design of the system (hereinafter referred to design) are the key stages in the process of its creation. Properly conducted, they can significantly reduce the time of software construction, minimize errors, increase reliability, and simplify testing and servicing of the created system. Over the last years, system designing has been in the centre of interest among specialists in this field. As a result, many methods that allow for systematic and analytic approach to software design have been worked out. JSP is one of them.

The JSP method (Jackson Structured Programming further called Jackson method) is a well documented and proven method of software design. It is completely independent of the language used. The method was created in 1976 by Michael Jackson and has become a widely used method of design, especially in Europe. In the 70s and 80s of the 20th century it was considered the standard software specification by WHO and the government of the United Kingdom [8]. This method is most frequently used to deal with sequential problems, particularly when the sequence is arranged in time.

The basic principle of Jackson method is that the design starts with the analysis and modeling of part of reality in the context of which the system is to work and which is to include. However, no data structures or methods that the system has to perform are specified. The system created with JSP method contains direct simulation (model) of reality which has to be specified prior to any designing [5].

Creating a model of reality is generally easier than taking decisions in the design stage. This is because the model frequently exists in reality and is well known to potential users of the system. When modeling a real problem, a software developer should grasp the user's point of view on that problem. Consulting the user may help avoid confusion and many errors in the future. A clear picture of the model defines possible methods and objects of the proposed system.

Another principle of Jackson method is that an appropriate model of sequential problem is also sequential. The model consists of a sequence of processes that communicate with one another. The sequence should reflect sequentiality of the actual problem. The model should be implemented through specification transformation into an efficient and useful set of processes, adapted to the available hardware. Special attention should be paid to the correct schedule of the processes, and particularly to the fact that a potentially small number of available processor units have to be shared by relatively many separate processes.

The Jackson method-based design process consists of three stages:

– Analysis
– Specification
– Implementation

The first two stages are responsible for the design of appropriate models of the system. The first stage is responsible for creating the data structure. In the second stage, system specification model and program structure model are created. The final stage means encoding of the existing models using a chosen programming language. Sometimes system implementation models are also created; this, however, will not be described here.

Data structure model construction is of major significance in Jackson method. The design of this model outstrips all other designs, according to the assumption that it is the data structure model that defines the potential space of the system function models. This is also consistent with the basic design principle to treat the system as a pattern transforming input data into output data, with very restrictive assumptions as to the form of data structures. These structures can only be represented as a hierarchical tree and are presented graphically as rectangles which can be interconnected by a relationship denoting inclusion. There are three basic types of the relationship: sequence, selection and iteration. They have been presented in Figure 1.

The sequence (Figure 1a) means that the structure A consists of components B, C and D. The choice (Figure 1b) means that the structure A consists of one of the components B, C or D. The iteration (Figure 1c) means that the structure A potentially contains multiple components C (may not include any of them). It should be emphasized that these basic types of tree-like structures can be mixed with one another to fully reflect the structure complexity. Additionally, each component structure may itself contain multiple smaller parts. This is a consequence of the decomposition rule mentioned above. Having developed the full data structure, the application structure model and system specification model should be designed. At this stage, it is assumed that the program structure model inherits from data structure model [6] [7]. This assumption is usually met. If the information consists of separate elements, the procedure processing this information can be divided into appropriate subroutines that process the respective elements. Likewise, processing an alternative means choosing a specific subroutine depending on which component has to be processed. The iterative procedure consists of multiple calls to the same subroutine for each of the components. The appropriate diagrams look exactly the same as in Figure 1, except that they no longer represent data structures, but the program structure that processes them.

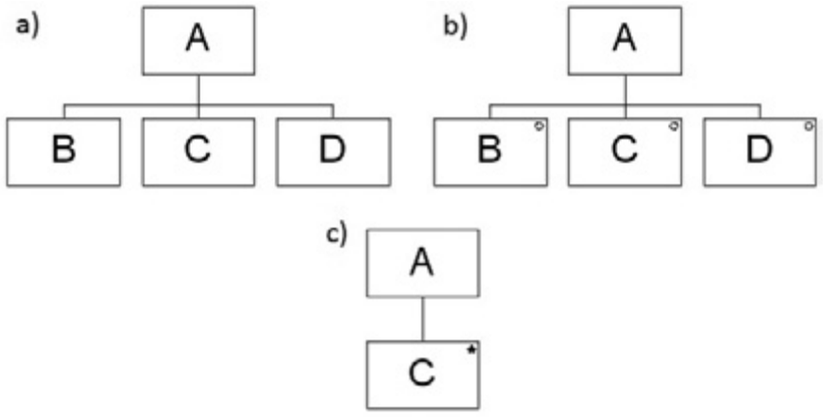


Fig. 1. Tree construction model: (a) sequence, (b) choice, (c) iteration.

The Jackson method perfectly suits to design a complex system involving a number of interrelated modules. An example of such a system is the application used to support management of education and research activity of a clinical department. The system operates on a variety of datasets, performing many operations specific to a particular scenario. The scenarios in this case depend on the posts occupied by the system users. Having in mind the enormous complexity of the system logic it is necessary to prepare an earlier detailed documentation in order to avoid complications at later stages of its construction.

To apply the method to design the system that supports management of a clinical department, the following datasets should be specified first [3]:

– all data (entities) inputted to the system (documents that are entered to the system),
– all data (entities) outputted from the system (documents that are presented to the user by the system),
– set of auxiliary data (entity) (used internally as system information store).

Complex information system is split into modules prepared for individual scenarios of system use. The scenarios are related to the posts at which the system is used and so is the list of data ranges used by the information system that supports the clinical department. General data can be divided into 4 categories:

– administrative and financial,
– scientific activity,
– didactic activity,
– medical.

In everyday work of the information system, the following posts can be listed:

– the Secretary's Office
– the Head of Department
– Teacher
– Scientific worker
– Doctor
– Student

Now, we are going to construct fragments of the input structure diagram for the Teacher module. Most frequently data used at this post include:

– teaching room occupation – a list of available rooms together with timetable related to them. Complete data on room reservation by respective teachers,
– list of students – full listing of students attending classes, along with data concerning the subject, course and year of study,
– examination and credit protocols – electronic version of protocols transmitted to the Dean's Offices with information concerning evaluation of individual students,
– distribution of students into groups – information about groups in which students take courses and specifying group membership for each student,
– students' attendance lists – electronic list of students attending each class,
– tests, credits and examinations (marks) – a collection of data required to make a decision whether to award a credit to a student or not,
– class synopsis – information about the respective syllabuses.

Data most frequently entered include:

– students' attendance,
– test, credit, exam results,
– exam and credit protocol data,
– class synopsis

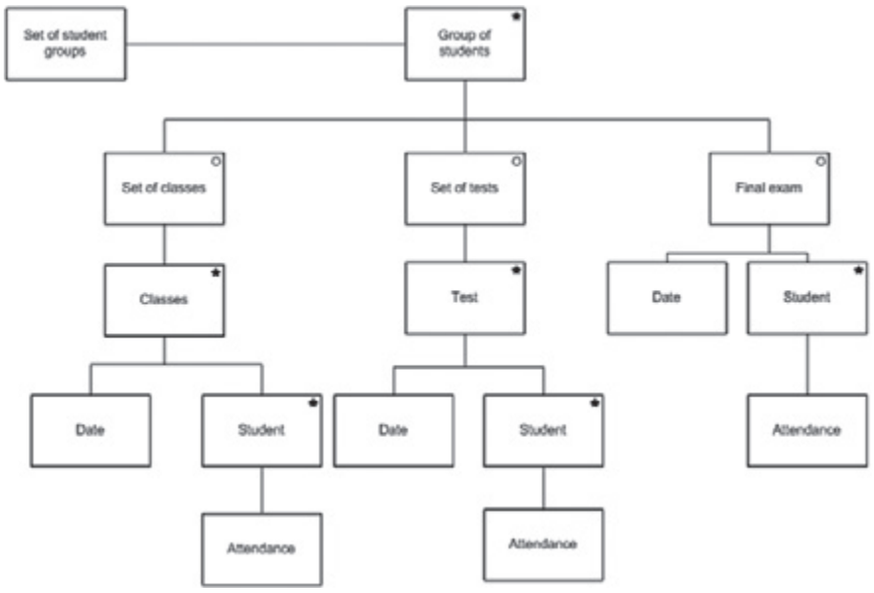


Fig. 2. Class attendance information data structure diagram.

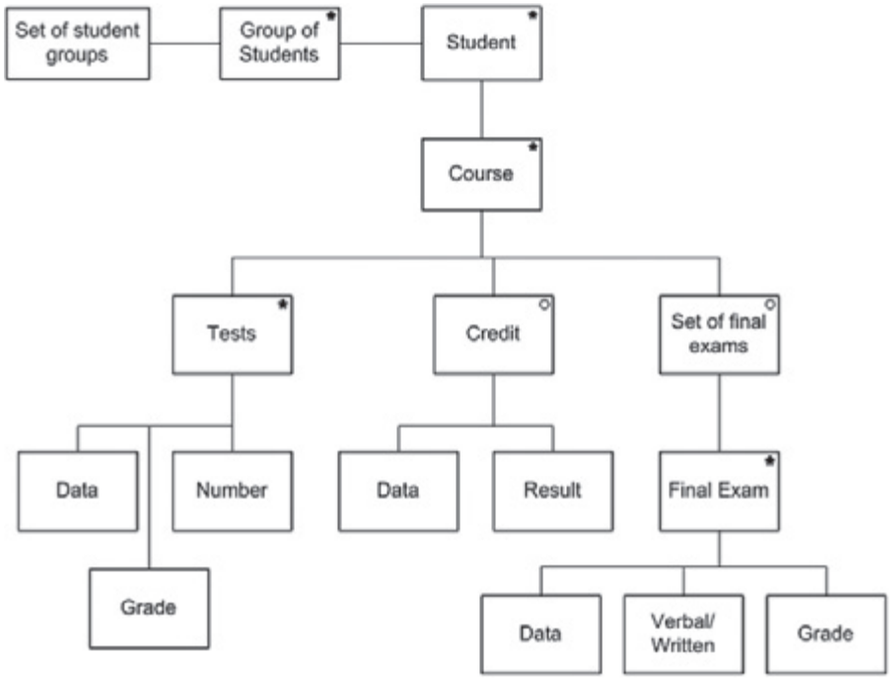


Fig. 3. Diagram of data structure information about students' evaluation results.

Checking students' attendance is one of the most frequent actions performed by teachers. The attendance is usually recorded in files, often in the form of sheets of paper cards with information about absence and late comings. This information has a specified format. Figure 2 shows the format in a JSP data structure diagram.

Checking attendance starts with choosing the respective student group. Next step involves choosing the type of meeting, based on the synopsis entered earlier. There are four options (regular classes, a test or exam). In the case of regular classes, the class subject should be found from the synopsis. Similarly, the previously planned scenario can be chosen for the test. Then, the date of the meeting should be entered and attendance has to be checked for each student. There are four alternative student attendance states: present, absent, late, excusably absent.

Verification is an essential element of the didactic process. The system supporting the process of teaching should therefore enable the user to record data concerning the results of student evaluation. Notes that are often made on separate sheets of paper or in specially designed notebooks can be transferred to the system, thus allowing easy access, not only for the teacher but also for all other authorized users [9]. In addition, incorporation of appropriate logic into the system will allow effective

creation of electronic versions of test or examination protocols. The structure that stores this information can be described by the diagram shown in Figure 3.

Entering the evaluation results begins with student group selection. For each student, course specification and type of verification (test, credit, exam) should be defined. Worthy of note is that for any combination of one student and the selected course it is possible to enter results of several tests (or none), and alternatively exam or credit grade. In the case of test, one of the scheduled tests has to be chosen. Any additional data, such as date and whether it is a resit, should be specified. In addition, student's grade, obtained during verification should be entered. For credit, date should be also specified, as well as the result (positive or negative). In the case of exam, two types can be entered – oral or written. The structure allows us to enter the data concerning the two types of examination at once, as well as the examination diet (first attempt or resit) by using the "Exam Data".

The examination and credit protocols are used periodically in a teaching unit. They are produced at the end of each semester, for each group of students and course. It contains data on final credits or exams and eventually has to been passed to the Dean's Office. Data structure of the examination and credit protocols is presented in the diagram (Figure 4).

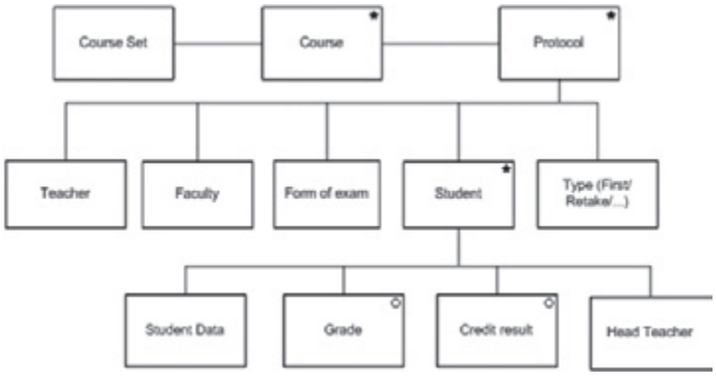


Fig. 4. Credit and exam protocol data structure diagram.

Such a protocol is designed for a specific course, selected from a group of courses. Thus, each course requires a separate protocol or even a few protocols (e.g.: first attempt and resit). In one instance of the protocol, additional data can be added about teacher, field of study or form of final testing (credit or exam). The main component of the protocol is a list of students with their data (full name or index number) and final test results. Credit result is either positive or negative, whereas examination result is expressed by a grade. In addition, each student must be assigned to the teacher who authorizes the evaluation results.

The initial stage of the teacher' coursework in the relevant subject involves designing the class synopsis, containing a detailed specification of the subject matter for each meeting with the group, together with details of the issues covered. The program content, carefully assigned to each meeting, is a key to efficient realization of the teaching process. The electronic version of the synopsis entered to the system will facilitate control of the teaching process, both in individual and group aspects. Relating the subject matter specified in the synopsis to the respective classes for students provides the information on the course program realization. The diagram of the synopsis data structure has been presented in Figure 5.

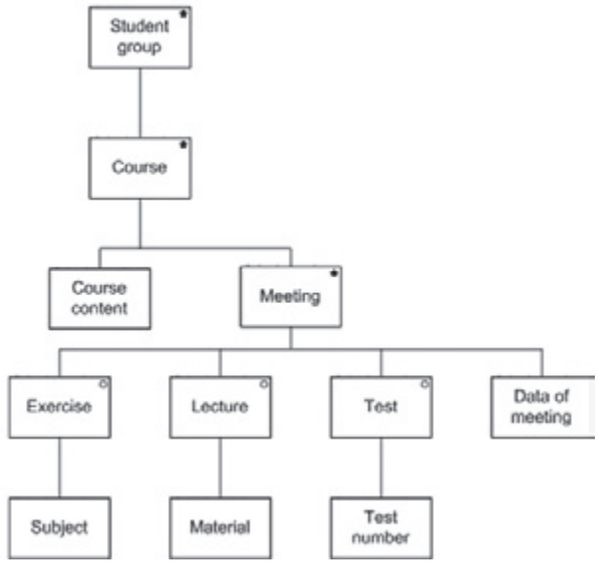


Fig. 5. Synopsis data structure diagram.

Synopsis shall be prepared separately for each group of students, taking into account the distribution of group meetings (their number and duration). It seems obvious that once the group has been defined, the course synopsis should be selected. The synopsis consists of certain number of meetings, each of which having a definite type: classes, lecture or testing. Additional pieces of information, such as duration or required teaching aids should be determined. Depending on the type of meeting we can specify task contents, material to discuss or a subsequent test number.

The computer system based on the JSP method ensures compliance with the assumptions made at the stage of analysis and allows tracking its further efficient expansion. Organization of input data structures allows their smooth conversion into elements of the application code responsible for their processing.

The complexity of the system that supports management of a clinical, teaching and research department makes it impossible to draw up the entire diagram in the current publication. However, the presented fragments prove that with proper effort the presentation of the complete diagram can be feasible.

The elements of the data structure diagram presented above constitute only a small part of the whole project. After correct analysis and decomposition of the problem, the diagram can be completed with new items. Such necessity may occur during discussing the system functionality with its future users. In such situations, it frequently appears that additional data can increase work efficiency, which should be one of the key benefits of the system implementation [2]. Definition of the diagram for the output and auxiliary data will clarify the tasks of the processes mediating data transformation.

References

1. Adamczewski P.: Zintegrowane systemy informatyczne w praktyce. Wydawnictwo MIKOM: Warszawa 1998.
2. Bennett C.J.: Computers, personal data and theories of technology: Comparative approaches to privacy protection in 1990s. *Science, Technology & Human Values* 1991, vol. 16, no. 1: 51-69.
3. Beynon-Davies P.: Inżynieria systemów informacyjnych. Wprowadzenie. Wydawnictwa Naukowo-Techniczne: Warszawa 2004.
4. Flasiński M.: Wstęp do analitycznych metod projektowania systemów informatycznych. Wydawnictwa Naukowo-Techniczne: Warszawa 1997.
5. Jackson M.A.: A System Development Method. Tools and Notions for Program Construction, Neel D. (ed.). Cambridge University Press 1982, pp. 1-26.
6. Jackson M.A.: Principles of Program Design. Academic Press 1975.
7. Jackson M.A.: JSP in Perspective. In: Broy M., Denert E. (eds), Software Pioneers: Contributions to Software Engineering. Springer 2002.
8. Ourusoff N.: Using Jackson Structured Programming (JSP) and Jackson Workbench to Teach Program Design. *Informing Science*, June 2003.
9. Płodzień J.: Stemposz E., Analiza i projektowanie systemów informatycznych. Polsko-Japońska Wyższa Szkoła Technik Komputerowych 2003.

INSTRUCTION FOR AUTHORS

The scope

The edition of the journal *Bio-Algorithms and Med-Systems* was initiated to open the discussion in the interdisciplinary scientific activity focused on medicine in both: experimental and theoretical research including the computer science application to diagnostics, therapy and education. The molecular level oriented problems (*Bio-Algorithms*) as well as technical applications to the hospital use (*Med-Systems*) are in the focus of our interest. The journal is open to cover variety of interests.

The journal is edited quarterly by *Jagiellonian University, Medical College,* Kraków, Poland.

Each corresponding Author is also asked to give information about potential reviewers .

Manuscript should be prepared in English. Usage of correct language is the responsibility of the Author.

Journal *Bio-Algorithms and Med-Systems* is registered in SCOPUS database monitoring citations of publications. The procedures of registration in other databases have been undertaken.

Each submitted material (manuscript) undergoes the peer review by two reviewers. Author is asked to answer to the reviewers comments and/or follow the advice correcting the paper according to reviewers' suggestion.

Polish Ministry of Science and Education journal rating: 6.000.

Submission of manuscript to *BAMS* implies that the copyright of the entire material except those already copyrighted by other publications like tables, figures etc. and those permission the Author(s) have already obtained, belongs to the Publisher.

The text shall be prepared in *.docx (preferable) or *.doc format with good quality of graphic of figures.

The article organization:

1. Title, Authors names, affiliation, key words (max 6)
2. Abstract – no more than 250 words
3. Introduction – describing the Interests and the main idea of the paper in relation to other publications introducing the Reader to the current Known How
4. Materials and Methods – the description of data sources, tools applied for the work
5. Results – presentation of the new original achievements
6. Conclusions/Discussion – presentation of the further plans, the relation to previous works to make possible the comparison of the results with others. The critical comments are also expected.
7. Acknowledgements – the source of founding and information as to the personal collaboration and discussion with specialists not included in Authors' list
8. References: the last section of the paper presenting the references in order of their appearance in text in form [1,2].
   Examples:
   [1] Pyrczak W, Sarapata K: Instructions …, Bio-Algorithms and Med-Systems 1, 386-387, 2005.
   [2] Tadeusiewicz R, Ogiela M. R. Medical Image …, Springer Verlag, Heidelberg, 2004
9. Tables and Figures should be numbered consecutively using Arabic numbers.
10. Color Figures – Cost of printing color figures will be charged to the Authors at the rate of U.S. $ 125 per plate of four color figures.
11. Reprints – in order for reprints should be placed at the time of submission of the article itself, so that it may be out for printing without loss of time, if found suitable for order for not less than 100 reprints is to be made.

MANUSCRIPT SUBMISSION

The only way to submit the paper possible is through the e-mail address *bams@cm-uj.krakow.pl.*

The electronic service *submission_office* (in preparation) will be available soon on the webpage www.bams.cm-uj.krakow.pl

The corresponding Author will/may be asked to submit:

1. cover letter presenting the information usable for the Editorial Board explaining the accordance of the paper with the specialization field represented by the Journal.
2. list of potential reviewers (together with the email addresses)
3. the main body paper
4. figures (tables shall be put into the main body)
5. ethic and copy right transfer questions will be asked