Spectral analysis package EPRSIM-C

Janez Štrancar

Laboratory of Biophysics – EPR center, Jožef Stefan Institute, SI-1000 Ljubljana, Slovenia

Introduction

EPR spectroscopy in combination with spin labeling proved to be one of the most desirable techniques in exploration of biological system partially due to sensitivity to local molecular motion and other properties and partially due to the fact that EPR is a nondestructive method enabling one to study complex system without splitting it into building blocks.

Although there was substantial development in spectral simulations during last decades the users often face to a problem of being overwhelmed by numerous spectral parameters which arise on one side due to precise spectral simulation models and on the other side due to complexity that should be taken into account when exploring biological systems. The EPR simulation software packages that were available in this research field paid insufficient attention to the problem of extraction of such a number of spectral parameters – the optimization routine were not chosen according to the complexity of the phase space and inverse problem appeared to be ill-posed many times due to strong numerical correlations in parameter space.

Therefore we develop EPRSIM software package that recently evolve into more general EPRSIM-C software package to be capable of doing automatic nitroxide-based characterization of biological system by implementing simpler and still accurate simulation models, powerful optimization routines as well as data condensation techniques that enable user to parameterize the complex system in a more understandable way. Many ideas of the whole approach can be found in the chapter *"Advanced ESR spectroscopy in membrane biophysics"*

EPRSIM-C software package provides **a tool for nitroxide-based characterization** of:
- specifically and nonspecifically labeled biological membranes,
- site directed spin labeled proteins,
- biopolymer networks explored by specific labeling or concentration imaging,
- nanomaterials and
- other interesting complex biological systems labeled with different labeling techniques.

EPRSIM-C software package simulates nitroxide EPR spectra as well as solves inverse problem involved in characterization procedure. For increased efficiency the different tasks are handled within different programs.

EPR experiments

EPR-Convert
• converting formats

EPRSIM-WIZ
• arranging datasets
• defining simulation models

EPRSIM-BBW
• dHEO / SD optimization
• run in background with parallel threads

GHOSTMAKER
• condense and filter the optimized solutions
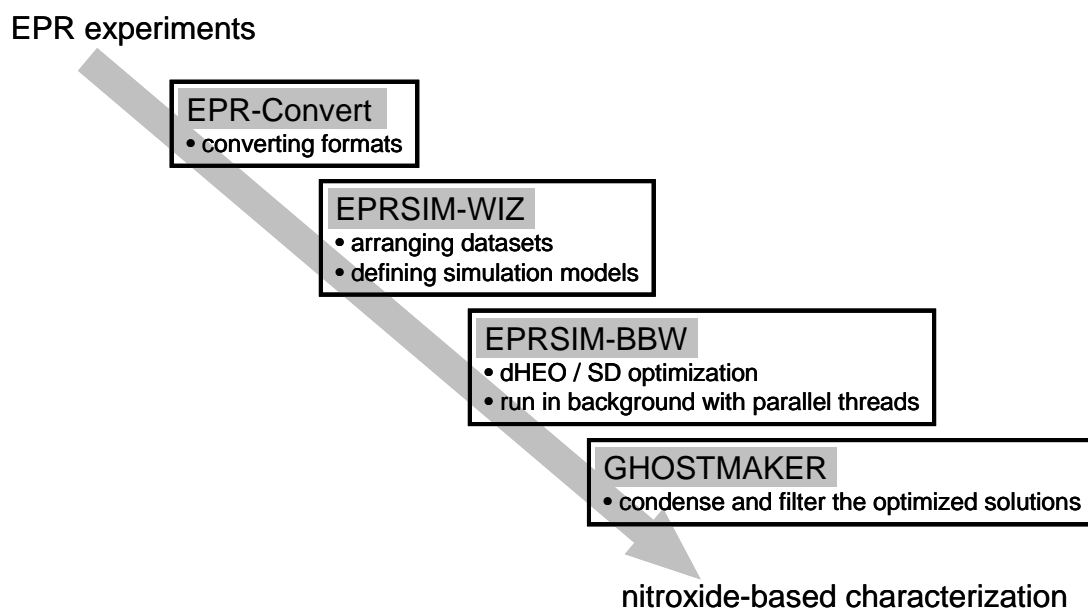
nitroxide-based characterization

Figure 1: Illustration of the EPRSIM-C software package concept with purposes of individual programs

All the intensive numerical tasks – optimizations – are done in background within **EPRSIM-BBW** in the way that one can submit the tasks from the network in a simple way. The simulation models are based on motional partial averaging as described in the chapter "*Advanced ESR spectroscopy in membrane biophysics*". The user must understand that the complex phase space can only be search by a stochastic population-based optimization routine which takes lots of computational time. Much faster local search methods are not applicable in general as the user can never verify if there are at least equally good solutions in other parts of the phase space.

The user interface of **EPRSIM-WIZ** helps to arrange and check the experimental and simulated datasets and is not meant to show the progress of the optimization routines. The main idea of having this kind of interface separated is to arrange clearly the experimental and simulation data. Various formats of experimental series can be read directly or translated into proper format within **EPR-Convert**.

Inverse problem solving is separated from data condensation routine implemented in **GHOSTMAKER** to increase the efficiency of calculations. This last program filters the datasets that came from optimization routine according to the goodness of fit and solution density. In addition it help to condensate the solutions into groups of various complexity.

List of abbreviations

BBW ...   EPRSIM-BBW, program that execute the optimization routines
WIZ ...   EPRSIM-WIZ, user interface program to present and arrange datasets
MTP ...   MTP filename with experimental and simulation paramteres of multiple
           datasets

SPC … OS-9 spectral file format
DTA … Elexsys spectral file format
ISO … 1st simulation model: isotropic tumbling
MEM … 2nd simulation model: anisotropic tumbling with full averaging over long axes
LLE … 3rd simulation model: isotropic spin-exchange label-label
LBE … 4th simulation model: isotropic spin-exchange label-broadening agent
MES … 5th simulation model: anisotropic tumbling with partial averaging of all rotations
TRP … 9th simulation model: spin trap simulations
GA … genetic algorithm optimization method
DSO … Downhill Simplex local search optimization method
HEO … hybrid evolutionary optimization method, sometimes HEA can also be found in literature; it represent GA hybridized with DSO
dHEO … HEO with shaking operator implemented
KBO … knowledge based operators
GHOST… crossection diagram to represent solutions after GHOST condensations using RGB color coding

Main characteristics

Spectral simulation modules use various models applicable for biological system in their native environment and physiological conditions. Heterogeneity principle is one of the most important properties of the approach. The other one is the implementation of fast motion approximation with partial averaging and geometry consideration that make this a tool, which can be used on any personal computer. To get the idea of computational demand of all the steps in the characterization procedure, we provide here a list of time intervals typical for current desktop processors with computing power of few GFLOPS:

- 1-10 ms for simulation model,
- 1-10 s for one local search,
- 1000 s for one stochastic population-based search
- 10 h for completed characterization procedure.

The user can easily scale the time demand for characterization procedure if the time for simulation model is increased dramatically. In addition this time demand increases dramatically when the inverse problem becomes close to ill-posed, i.e. when the number of spectral parameters is increased in such a way that strong correlations appear in the parameter space.

Autonomous optimization routine (implemented in **EPRSIM-BBW**) derived by hybridization of evolutionary optimization methods and local search routine handle inverse problem solving with large number of spectral parameters involved. In addition many knowledge-based operators were introduced to improve numerical consistence of the simulation models. Multiple optimizations are needed as the optimization routine belongs to the class of stochastic algorithms. Since the inverse problem solving routine requires large computing power, the **EPRSIM-BBW** can do multiple optimizations in parallel threads making it more efficient on multi-processor

platforms. **EPRSIM-BBW** runs in background and continuously scans the desired directory for tasks to be performed.

User interface is settled-up within **EPRSIM-WIZ** that allows data handling and simulation presentation for various kind of EPR experiments and models of different complexity like. Multi-frequency experiments, multi-angle experiments as well as other experimental series can be analyzed. Several formats can be read directly or translated with **EPR-Convert**.

As the multiple optimization routines provide huge amount of datasets, one has to find which solutions are most significant and reliable. This is done by GHOST condensation algorithm implemented in **GHOSTMAKER**. The populations of solutions are compared to each other for goodness of fit, passing through only the solutions that are best. In addition, there is density filtering, that pass only those solutions that has many other solutions in their neighborhood (in the parameter space). Finally, the solutions are grouped, i.e. condensed to make parameterization more transparent.

EPRSIM-C library

The general library of EPRSIM-C programs contains all the modules that are shared among the EPRSIM-C programs, like handling with spectral and MTP files, simulation models and optimization routines.

MTP file structure

The name "MTP" abbreviates multiple datasets. The MTP file is divided into two parts: system constants (block A) and spectral parameters (block B). As it is a text file it can be easily modified manually without using EPRSIM-WIZ if desired.

The system part (block A) includes:
- the optimization constants like different sizes, probabilities, switches, etc.
- the simulations models' constants like limits, uncertainties, etc.
- nitroxide tensors' elements
- artificial noice value
- number of optimization runs (i.e. "copies") that should be made to achieve appropriate statistics

The following spectral parameters' part (block B), the data structure must be provided for each of the experimental datasets:
- number of experiments that were performed on the same sample and should be simulated by the same spectral simulation parameters' group (like in multi-frequency experiments)
- filenames of the experimental datasets
- experimental constrains: center field, sweep width, frequency and angle

- number of spectral components (up to 5)
- spectral parameters, error and switches (for the optimization) for each of the spectral components including spectral simulation model that should be applied for each particular spectral component
- type of measure of goodness of fit

For the precise structure see the implementation notes and manuals of the EPRSIM-C package.

When opening spectral datasets, the following implementations should be noted:
- additional Gaussian noise is added to experimental spectral data with the amplitude as defined in MTP file,
- standard deviation of the noise is defined by fitting the outer wings ("zero-signal") of the spectrum by two quadratic functions,
- rough baseline correction is done according to the derivative of the 3rd degree polynomial fit to the outer wings of the 1st integral of the spectrum,
- rough correction of the centre-field position is done by aligning the maximum of the first integral of the simulated spectrum to the maximum of the first integral of the experimental spectrum,
- fine correction of the amplitude (normalization), centre-field and baseline are done together with spectral simulations in simulation module to adapt also to long wings of the spectra and high noise amplitudes.

The simulation models

One of the most important information saved in EPRSIM library is the structure of **the simulation models**. The following simulation models (with three letter abbreviation) can be used:
- 1st model: isotropic tumbling (ISO)
- 2nd model: anisotropic tumbling with full averaging over long axes (MEM)
- 3rd model: isotropic spin-exchange label-label (LLE)
- 4th model: isotropic spin-exchange label-broadening agent (LBE)
- 5th model: anisotropic tumbling with partial averaging of all rotations (MES)
- 9th model: spin trap simulations (TRP)

Models from 6th to 8th position are only for developing purposes and should not be used routinely. For physical background that is not explicitly discussed here see the chapter "*Advanced ESR spectroscopy in membrane biophysics*".

The following general principles are used in the implementation of all the simulation models:
- weighting of the spectral components is done according to double integrals (spectral intensities)
- each spectral parameter has a "switch" which defines whether this particular parameter will be optimized within optimization routine (see for example Table 1); note, that some models contains spectral parameters there are not meant to be optimized in general as they represent some variations of the models, therefore they have switch setup to false by default
- linear baseline correction is done together with the amplitude normalization after weighting and composing of the spectral components

- goodness of fit is calculated after normalization and baseline correction according to the χ^2 type (defined in block B of MTP file for each experimental dataset) – $\beta=0$ for standard χ^2 and $\beta=1$ for first order island-weighted χ^2

$$\chi^2_{(\beta)} = \frac{1}{N} \sum_{island} \left((\lambda_{island})^\beta \sum_{i=1}^{\lambda_{island}} \frac{(y_{exp} - y_{sim})^2}{\sigma^2} \right)$$

1st model - isotropic tumbling (ISO)

The ISO simulation model can be used to simulate EPR spectral component of isotropic tumbling nitroxide spin label in a fast motional regime.

Table 1: Parameter space of the 1st model - isotropic tumbling (ISO)

Name of the parameter	Notation	Unit	Definition interval	Switch
proportion of Lorentzian/Gaussian mixture	L_w		0..1	T
isotropic rotational correlation time	τ_c	ns	0.02 .. 3	T
additional broadening constant	W	G	0.1 .. 4	T
polarity correction p_A	p_A		0.8 .. 1.2	T
polarity correction p_g	p_g		0.9996 .. 1.0002	T
Weight	D		0 .. 1	T

Here are some implementation notes about ISO model:
- Fast motion approximation determines part of the linewidth,
- W is added to the linewidth of all lines directly

$$\frac{1}{T_2} = A\left(\tau_c, \underline{A}, \underline{g}\right) + B\left(\tau_c, \underline{A}, \underline{g}\right) M + C\left(\tau_c, \underline{A}, \underline{g}\right) M^2 + W$$

- Mixture of Lorentzian/Gaussian lineshape is calculated at the resonant line positions; $L_w=0$ for pure Gaussian line
- Polarity correction factor are used to correct the scalar of the tensors (trace-scalar correction)
- Hyperfine structure is added for C^{13} with the abundance determined in the MTP file (section "Tensors:")

2nd model - anisotropic tumbling with full averaging over long axes (MEM)

The MEM simulation model can be used to simulate EPR spectral component of nitroxide spin probes that are fastly tumbling in an anisotropic environment like membrane.

Note, that the rotations around long-axes are assumed to be so fast to allow full averaging of the magnetic tensor values over this direction. The anisotropy of the rotational motion is therefore described by one order parameter Szz.

Note that the model implements only one efficient rotational correlation time which is valid to describe the wobbling of the long axes.

Table 2: Parameter space of the 2nd model - anisotropic tumbling with full averaging over long axes (MEM)

Name of the parameter	Notation	Unit	Definition interval	Switch
order parameter S$_{zz}$	S		0..1	T
isotropic rotational correlation time	τ_c	ns	0.02 .. 3	T
additional broadening constant	W	G	0.1 .. 4	T
Polarity correction p$_A$	p$_A$		0.8 .. 1.2	T
Polarity correction p$_g$	p$_g$		0.9996 .. 1.0002	T
Weight	D		0 .. 1	T

Here are some implementation notes about MEM model:
- Fast motion approximation determines part of the linewidth,
- W is added to the linewidth of all lines directly (see model 1)
- Polarity correction factor are used to correct all the tensor components by the same amount (trace-scalar correction)
- S$_{zz}$ order parameter is used to determine partial averaging of the tensor component due to wobbling of the long molecular axes
- Complete averaging is used for motions around the long axes
- Stick spectrum is generated by the numerical computation with the $\pi/2$ interval divided into 2000 steps by taking into account also transition probability correction factors up to second order
- Second order correction to stick spectrum is made near singularities by analytical expansions
- Renormalization is done to stick spectrum after all corrections
- Convolution of pure Lorentzian derivative and stick spectrum is done up to maximal ten linewidths from each point

3rd model - isotropic spin-exchange label-label (LLE)

The LLE simulation model can be used to simulate EPR spectral component of concentrated nitroxide spin probes in isotropic environment. Therefore label-label spin exchange is expected.

Table 3: Parameter space of the 3rd model - isotropic spin-exchange label-label (LLE)

Name of the parameter	Notation	Unit	Definition interval	Switch
isotropic rotational correlation time	τ_c	ns	0.02 .. 3	T
additional broadening const. (not exchange)	W	G	0.1 .. 4	T
Spin exchange rate	W$_{ex}$	G	0.1 .. 10	T
polarity correction p$_A$	p$_A$		0.8 .. 1.2	T
polarity correction p$_g$	p$_g$		0.9996 .. 1.0002	T
weight	d		0 .. 1	T

Here are some implementation notes about LLE model:
- linewidth is determined as in model 1 using τ_c and W

- Spin exchange rate Wex is used in Bloch equation to determine lineshapes for spin-spin interacting system with the use of complex magnetization G

$$\left(\frac{1}{T_2(M_1)} + i\left(B - B_{M_1}\right)\right)\mathcal{G}_{M_1} - \sum_{M_1 \neq M_2} W_{M_1 M_2}\left(\mathcal{G}_{M_2} - \mathcal{G}_{M_1}\right) = -iB_1 \mathcal{M}_0$$

- Polarity correction factors are used as in model 1

4th model - isotropic spin-exchange label-broadening agent (LBE)

The LBE simulation model can be used to simulate EPR spectral component of nitroxide spin probes in isotropic environment with high concentration of broadening agents. Therefore label-broadening agent spin exchange is expected.

Table 4: Parameter space of the 4th model - isotropic spin-exchange label-broadening agent (LBE)

Name of the parameter	Notation	Unit	Definition interval	Switch
isotropic rotational correlation time	τ_c	ns	0.02 .. 3	T
intrinsic linewidth of broadening agent	W_{Br}	G	100 .. 3000	T
concentration of broadening agent	cBr	mmol/l	100 .. 10000	T
polarity correction p_A	p_A		0.8 .. 1.2	T
polarity correction p_g	p_g		0.9996 .. 1.0002	T
weight	D		0 .. 1	T

For the implementation notes see LLE model.

5th model - anisotropic tumbling with partial averaging of all rotations (MES)

The MES simulation model can be used to simulate EPR spectral component of a nitroxide spin probes in anisotropic environment like membrane or of a nitroxide spin probes attached to protein (site directed spin labeling). It is a generalization of MEM model.

Instead of describing the anisotropy of rotational motion by a single parameter order parameter it has two cone angle parameters describing the wobble model (main cone angle and asymmetry angle) to generalize the restriction to rotational motion. The reason of implementing wobble model instead of Gaussian distribution is that one can easily translate the two cone angles into order parameters (assumed that the probability function is non-zero constant up to the two angles and zero above their values), however, it would not be possible to translate order parameters easily into cone angle which are more native choice in the case of site directed spin labeling.

In order not to introduce strong correlation into parameter space, the implementation of the second rotational correlation time is avoided.

Note that this model allows to vary the geometry of the orientation distribution of spin probes by geometrical parameter g_1 within the sample in EPR spectrometer: it can be isotropic, planar, cylindrical or pore/neck-like. The other geometrical parameter g_2 is for developing purposes only – do not change its value.

In addition, this model allows to smoothly vary the orientation of the tensor relative to the coordinate system of fast motional averaging by changing the two nitroxide rotation angles.

Table 5: Parameter space of the 5th model - anisotropic tumbling with partial averaging of all rotations (MES)

Name of the parameter	Notation	Unit	Definition interval	Switch
Geometrical parameter 1	g_1		0 ..	F
Geometrical parameter 2	g_2			F
Nitroxide rotation angle 1	th	rad	0 .. $\pi/2$	F
Nitroxide rotation angle 2	fi	rad	0 .. $\pi/2$	F
main cone angle	th_0	rad	0 .. $\pi/2$	T
asymmetry angle	fi_0	rad	0 .. $\pi/2$	T
effective rotational correlation time	τ_c	ns	0.02 .. 3	T
additional broadening constant	W	G	0.1 .. 4	T
polarity correction p_A	p_A		0.8 .. 1.2	T
proticity prot	prot		-0.07 .. 0.07	T
weight	d		0 .. 1	T

Here are some implementation notes about MES model:
- Polarity correction factor p_A used as a scalar to correct all the A tensor components
- Proticity correction factor prot used in the following way together with g_{fe} (free electron g), p_A and empirical constants:
$$g_{xx}:=g_{xx}+(g_{xx}-g_{fe})*(-0.77*(p_A-1)-prot)$$
$$g_{yy}:=g_{yy}+(g_{yy}-g_{fe})*(-0.06*(p_A-1))$$
$$g_{zz}:=g_{zz}+(g_{zz}-g_{fe})*(-0.12*(p_A-1))$$
- Nitroxide rotation angles 1 and 2 are used to get proper orientation of nitroxide coordinate system relative to molecular axes (axes of fast motion)
- Spectral lineshape is calculated using fast motion approximation by convolution of Lorentzian lineshape (linewidth defined in motional narrowing approximation using single effective rotational correlation time τ_c) with distribution of resonant lines (stick spectrum) which is calculated using partial averaging of the magnetic tensors' components and appropriate geometry of orientation distribution
- Although effective rotation correlation time τ_c is not allowed to exceed spectroscopy time window (t_{win}=3 ns), the linewidths t[j] (calculated from motional narrowing approximation) are corrected empirically for larger values based on linewidths tt[j] at t_{win}
$$t[j]:=tt[j]-(tt[j]-t[j])*Exp(-(\tau_c -t_{win}/3)/t_{win})$$
- The transformations between coordinate systems used in the algorithm bases on the following angles:
  - membrane-normal orientation dependence {Θ, Φ} and

- orientation dependence of fastly changing rotational conformations defined with the Euler angles $\{\psi, \vartheta, \varphi\}$

Note that in the software "text-like" notation is used:

- {psi, th, fi} for $\{\psi, \vartheta, \varphi\}$
- {TH, FI} for $\{\Theta, \Phi\}$

Maximal wobble angles are denoted by the bottom-right index "0".

- Hamiltonian applied was:

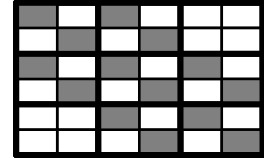$$\hat{H} = \mu_B B \; g^{\mathit{eff}}\left(\bar{\Omega}\right)\hat{S}_z + \hat{\vec{I}} \; A^{\mathit{eff}}\left(\bar{\Omega}\right)\hat{\vec{S}}$$

where partial averaging is taken into account with effective expressions of g and A tensor components but magnetic field orientation dependence remains explicit in vector $\bar{\Omega} = \{\Theta, \Phi\}$

- The most important averages of trigonometric functions of ϑ_0, φ_0, and ψ_0, which is approximated to be the same as φ_0, involves the following expressions:

$$\overline{\cos(\vartheta)^2} = \tfrac{1}{3}\left(\cos^2(\vartheta_0) + \cos(\vartheta_0) + 1\right)$$

$$\overline{\sin(\varphi)^2} = \frac{\varphi_0 - \tfrac{1}{2}\sin(2\varphi_0)}{2\varphi_0} = \frac{1}{2}\left(1 - \frac{\sin(2\varphi_0)}{2\varphi_0}\right)$$

$$\overline{\sin(\psi)^2} = \frac{\psi_0 - \tfrac{1}{2}\sin(2\psi_0)}{2\psi_0} = \frac{1}{2}\left(1 - \frac{\sin(2\psi_0)}{2\psi_0}\right)$$

$$\overline{\sin(2\vartheta)} = \tfrac{2}{3}\left(\cos(\vartheta_0) + 1\right)\sin(\vartheta_0)$$

$$\overline{\cos(\psi)} = \frac{\sin(\psi_0)}{\psi_0}$$

- Hamiltonian matrix is solved in "S_z (I_z, I_+, I_-) terms only" approximation with relative accuracy 0.01%, absolute accuracy 0.01 mT (including diagonals of the super-block-diagonal and sub-block diagonal)

- Resonant line position are calculated from

$$\Delta E_{M_I} = h\nu \quad \rightarrow \quad B = \frac{h\nu - M_I\sqrt{\left(A_{zz}^{\mathit{eff}}\left(\Theta_i, \Phi_i\right)\right)^2 + 2\left|A_{xzyz}^{\mathit{eff}}\left(\Theta_i, \Phi_i\right)\right|^2}}{\mu_B \; g^{\mathit{eff}}\left(\Theta_i, \Phi_i\right)}$$

with additional expressions

$$A_{zz}^{\mathit{eff}}\left(\Theta_i, \Phi_i\right) = \left[\widetilde{A}_{xx}^{avg}\cos^2(\Phi_i)\sin^2(\Theta_i) + \widetilde{A}_{yy}^{avg}\sin^2(\Phi_i)\sin^2(\Theta_i) + \widetilde{A}_{zz}^{avg}\cos^2(\Theta_i)\right.$$
$$\left. + \widetilde{A}_{xz}^{avg}\cos(\Phi_i)\sin(2\Theta_i)\right]\left(g^{\mathit{eff}}\left(\Theta_i, \Phi_i\right)\right)^2$$
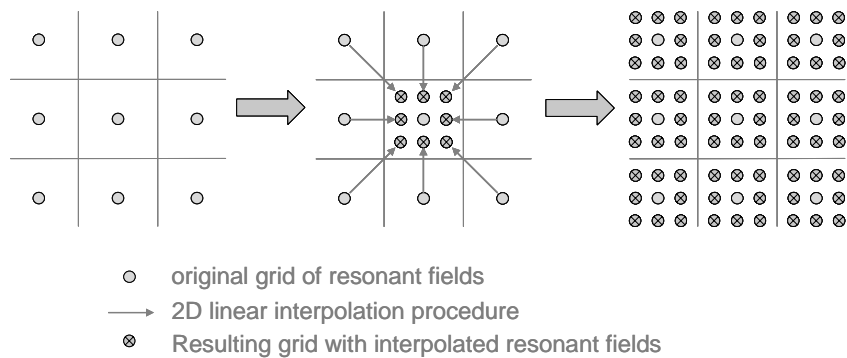
$$A_{xzyz}^{eff}(\Theta_i, \Phi_i) = \left[\frac{1}{2} \left(\frac{1}{2} \left(\tilde{A}_{xx}^{avg} \cos^2(\Phi_i) + \tilde{A}_{yy}^{avg} \sin^2(\Phi_i) - \tilde{A}_{zz}^{avg} \right) \sin(2\Theta_i) + \tilde{A}_{xz}^{avg} \cos(\Phi_i)\cos(2\Theta_i) \right) \right.$$

$$\left. + \frac{i}{2} \left(\frac{1}{2} \left(\tilde{A}_{xx}^{avg} - \tilde{A}_{yy}^{avg} \right) \sin(2\Phi_i)\sin(\Theta_i) + \tilde{A}_{xz}^{avg} \sin(\Phi_i)\cos(\Theta_i) \right) \right] \left(g^{eff}(\Theta_i, \Phi_i) \right)^2$$

$$\left(g^{eff}(\Theta_i, \Phi_i) \right)^2 = \tilde{G}_{xx}^{avg} \cos^2(\Phi_i)\sin^2(\Theta_i) + \tilde{G}_{yy}^{avg} \sin^2(\Phi_i)\sin^2(\Theta_i) + \tilde{G}_{zz}^{avg} \cos^2(\Theta_i)$$
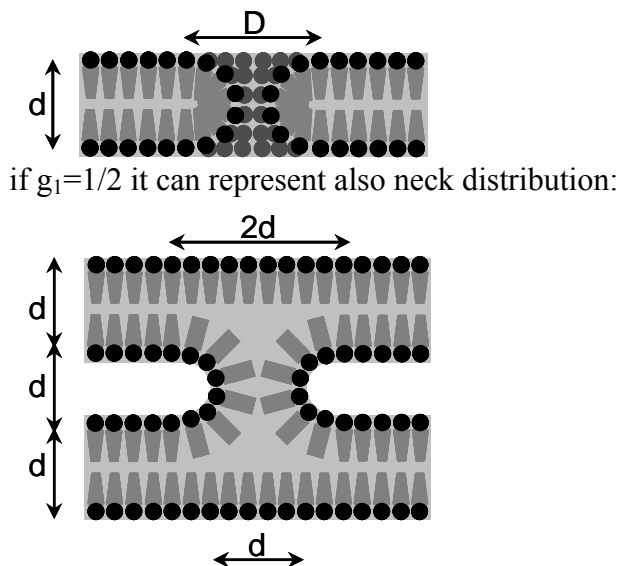
$$+ \tilde{G}_{xz}^{avg} \cos(\Phi_i)\sin(2\Theta_i)$$

and

$$\tilde{G}_{xx}^{avg} = g_{xx}^2 \left(\overline{\cos^2(\varphi)} \, \overline{\cos^2(\vartheta)} \, \overline{\cos^2(\psi)} + \overline{\sin^2(\varphi)} \, \overline{\sin^2(\psi)} \right)$$

$$+ g_{yy}^2 \left(\overline{\sin^2(\varphi)} \, \overline{\cos^2(\vartheta)} \, \overline{\cos^2(\psi)} + \overline{\cos^2(\varphi)} \, \overline{\sin^2(\psi)} \right) + g_{zz}^2 \, \overline{\sin^2(\vartheta)} \, \overline{\cos^2(\psi)}$$

$$\tilde{G}_{yy}^{avg} = g_{xx}^2 \left(\overline{\cos^2(\varphi)} \, \overline{\cos^2(\vartheta)} \, \overline{\sin^2(\psi)} + \overline{\sin^2(\varphi)} \, \overline{\cos^2(\psi)} \right)$$

$$+ g_{yy}^2 \left(\overline{\sin^2(\varphi)} \, \overline{\cos^2(\vartheta)} \, \overline{\sin^2(\psi)} + \overline{\cos^2(\varphi)} \, \overline{\cos^2(\psi)} \right) + g_{zz}^2 \, \overline{\sin^2(\vartheta)} \, \overline{\sin^2(\psi)}$$

$$\tilde{G}_{zz}^{avg} = g_{xx}^2 \, \overline{\cos^2(\varphi)} \, \overline{\sin^2(\vartheta)} + g_{yy}^2 \, \overline{\sin^2(\varphi)} \, \overline{\sin^2(\vartheta)} + g_{zz}^2 \, \overline{\cos^2(\vartheta)}$$

$$\tilde{G}_{xz}^{avg} = 0 \qquad\qquad \tilde{G}_{xy}^{avg} = 0 \qquad\qquad \tilde{G}_{yz}^{avg} = 0$$

$$\tilde{A}_{xx}^{avg} = \frac{A_{xx}}{g_{xx}^2} \left(\overline{\cos^2(\varphi)} \, \overline{\cos^2(\vartheta)} \, \overline{\cos^2(\psi)} + \overline{\sin^2(\varphi)} \, \overline{\sin^2(\psi)} \right)$$

$$+ \frac{A_{yy}}{g_{yy}^2} \left(\overline{\sin^2(\varphi)} \, \overline{\cos^2(\vartheta)} \, \overline{\cos^2(\psi)} + \overline{\cos^2(\varphi)} \, \overline{\sin^2(\psi)} \right) + \frac{A_{zz}}{g_{zz}^2} \, \overline{\sin^2(\vartheta)} \, \overline{\cos^2(\psi)}$$

$$\tilde{A}_{yy}^{avg} = \frac{A_{xx}}{g_{xx}^2} \left(\overline{\cos^2(\varphi)} \, \overline{\cos^2(\vartheta)} \, \overline{\sin^2(\psi)} + \overline{\sin^2(\varphi)} \, \overline{\cos^2(\psi)} \right)$$

$$+ \frac{A_{yy}}{g_{yy}^2} \left(\overline{\sin^2(\varphi)} \, \overline{\cos^2(\vartheta)} \, \overline{\sin^2(\psi)} + \overline{\cos^2(\varphi)} \, \overline{\cos^2(\psi)} \right) + \frac{A_{zz}}{g_{zz}^2} \, \overline{\sin^2(\vartheta)} \, \overline{\sin^2(\psi)}$$

$$\tilde{A}_{zz}^{avg} = \frac{A_{xx}}{g_{xx}^2} \, \overline{\cos^2(\varphi)} \, \overline{\sin^2(\vartheta)} + \frac{A_{yy}}{g_{yy}^2} \, \overline{\sin^2(\varphi)} \, \overline{\sin^2(\vartheta)} + \frac{A_{zz}}{g_{zz}^2} \, \overline{\cos^2(\vartheta)}$$

$$\tilde{A}_{xz}^{avg} = 0 \qquad\qquad \tilde{A}_{xy}^{avg} = 0 \qquad\qquad \tilde{A}_{yz}^{avg} = 0$$

- Stick spectrum is calculated based on the 2D-extrapolation of the master grid over Θ and Φ (orientation relative to magnetic field) where the numbers of master steps in Θ and Φ dimensions correspond to the maximal resonant field change for particular $\pi/2$ rotation at resolution of 0.6 G;

○   original grid of resonant fields
⟶   2D linear interpolation procedure
⊗   Resulting grid with interpolated resonant fields

- Geometrical parameter 1 (g_1) is used to define geometry of the orientation distribution relative to magnetic field:
  - $g_1 = -1$ is used to define isotropic distribution
  - $g_1 = 0$ is used to define planar distribution (perpendicular to magnetic field)
  - $g_1 = -2$ is used to define cylindrical distribution (symmetry axis parallel to magnetic field)
  - $0 < g_1 < 1$ is used to define pore distribution where g_1 represents the ratio between the bilayer thickness d and the pore outer diameter D:

if $g_1 = 1/2$ it can represent also neck distribution:

- Renormalization is done to stick spectrum after all corrections
- convolution of Lorentzian derivative and stick spectrum is done up to maximal ten linewidths from each point

9th model - spin trap simulations (TRP)

The TRP simulation model can be used to simulate EPR spectral component of a spin trap in isotropic environment or any other spin system interacting with up to 5 different kinds of nuclei regarding the hyperfine coupling.

Note that the spin values as well as number of spin are currently not meant for optimization (as they are not real numbers). However, regularly check the

EPRSIM –C support web page as the dHEO optimization routine is being modified to be able of optimization also in mixed continuous-discrete parameter spaces. In such a case the number of spins would also be possible to be optimized.

Table 6: Parameter space of the 9th model - spin trap simulations (TRP)

Name of the parameter	Notation	Unit	Definition interval	Switch
proportion of Lorentzian/Gaussian mixture	L_w		0..1	T
additional broadening constant	W	G	0.1 .. 4	T
polarity correction p_g	p_g		0.9996 .. 1.0002	T
spin of nucleus 1	S_1		Half integer: 0 .. 5	F
number of nucleus 1	N_1		Integer: 0 .. 10	F
hyperfine coupling with nucleus 1	a_1	G	0 .. 50	T
spin of nucleus 2	S_2		Half integer: 0 .. 5	F
number of nucleus 2	N_2		Integer: 0 .. 10	F
hyperfine coupling with nucleus 2	a_2	G	0 .. 50	T
spin of nucleus 3	S_3		Half integer: 0 .. 5	F
number of nucleus 3	N_3		Integer: 0 .. 10	F
hyperfine coupling with nucleus 3	a_3	G	0 .. 50	T
spin of nucleus 4	S_4		Half integer: 0 .. 5	F
number of nucleus 4	N_4		Integer: 0 .. 10	F
hyperfine coupling with nucleus 4	a_4	G	0 .. 50	T
spin of nucleus 5	S_5		Half integer: 0 .. 5	F
number of nucleus 5	N_5		Integer: 0 .. 10	F
hyperfine coupling with nucleus 5	a_5	G	0 .. 50	T
weight	d		0 .. 1	T

Here are some implementation notes about TRP model:
- Up to 5 different nuclei can be coupled to one radical (component)
- For each group (type of nuclei) multiplet positions and amplitude of lines is calculated using binomial coefficients and hyperfine coupling of the particular nuclei
- Convolution with mixture Lorentzian/Gaussian lineshape is implemented; L_w=0 for pure Gaussian line

Hybrid evolutionary optimization (HEA = GA hybridized with DS)

One of the most important issues of the EPRSIM-C package is the application of robust optimization scheme hybrid evolutionary optimization (HEO) based on generational genetic algorithm (GA). It bases on genetic-like transformations of parameters of a large population of M points, where M>> N_p, (each point is a set or vector of spectral parameters {p_i}), where N_p is the number of optimizing parameters. In our case GA is hybridized with local search operator (like DSO) and knowledge-based operators (KBO).

Starting population is created randomly within definition intervals [$p_{i,min}$, $p_{i,max}$].

Main loop consists of the selection, real coding and application of genetic operators in the following scheme (for Figure see the software implementation notes and Book chapter):
- fitness (χ^2) function evaluation for all new individuals
- sorting
- elitism implementation
- fitness-proportional or tournament selection
- genetic-operations application:
      - multi-point/uniform crossover
      - uniform mutation
      - local mutation (DSO)
      - knowledge-based operators
      - random replacement
The stop criteria for HEO used in EPRSIM C is maximum number of generations, which is 100.

Some special implementation notes are as follows:
- parameters' definition intervals are defined with minimal and maximal values in MTP file ("Limits:" section)
- grids defined in MTP file (section "Grids:") are used to define first step ("simplex")
- generation number is 50, 100, 100, 100 for up to 8, 8 - 14, 14 - 20 parameters and above, respectively
- population size is 130, 200, 280, 400 for up to 8, 8 - 14, 14 - 20 parameters and above, respectively
- number of crossover sites in multipoint crossover is 1, 1, 3 and 3 for up to 8, 8 - 14, 14 - 20 parameters and above, respectively
- other parameters (mutation probability, crossover probability, DSO probability, KBO probability, elite size) are defined in MTP file ("Hybrid Evolutionary Algorithm" section)
- elite cannot contain more than one copy of the same solution
- elite solutions replace the worst in the population every generation if the elite solutions cannot be found in the existing population (to prevent crowding of the same solution)
- tournament selection is used to determine the pool of solution that will be transformed by genetic operators (better solutions win for selection in random pairs)
- random selection is used for selection the pair of the solutions from the pool of tournament-selected solutions
- shaking is applied within the crossover with the shaking amplitude defined from product of Gaussian-randomly deviation (amplitude 0.5) and grids (defined in MTP file in "Grids:" section)
- knowledge based operator (KBO) 1 – $\tau_{c,1}$-d$_1$-$\tau_{c,2}$-d$_2$ equilibrator – multiplies smaller τ_c by 1.3 and larger by 0.7 and subtract half of the smaller weight from the weight of the component that has larger τ_c value and add the same amount to the other one;
- knowledge based operator (KBO) 2 – W_1-d$_1$- W_2-d$_2$ equilibrator – multiplies smaller W by 1.3 and larger by 0.7 and subtract half of the smaller weight from the weight of the component that has larger W value and add the same amount to the other one

- knowledge based operator (KBO) 2 – polarity sorter – remix the polarity correction factors p_g of the components in the opposite order of the sorted polarity correction factor p_A; note that polarity sorter should not be used for 5th model (MES)
- DSO is implemented as a mutation operator with probability defined in MTP file (section "Hybridization with DS during HEA")
- Resolution Auto tune adopt the grid for the (GA) mutation operator (initially defined by the grid constants from MTP file to standard deviations found by covariance matrix analysis every 10 generations
- after maximal generation number is achieved, substitution from elite is done ones more, population is sorted, single DSO is used on solutions of the better part of population as defined in MTP file (section "Hybridization with DS after HEA" – possible usage and part of population should be defined)

EPRSIM-C programs

EPRSIM-WIZ

The purpose of EPRSIM-WIZ program is primarily to handle EPR data, open spectral files, create multiple experiment in MTP files, collect or combining EPR experiments, save files (with corrected baseline, normalized, integrated), print the experimental datasets with simulations together, etc. Within EPRSIM-WIZ one usually defines the spectral simulation models that have to be used in optimization procedure for each of the experimental datasets. In addition, local optimization of spectral parameters by Downhill-Simplex optimization routine is possible for fine tuning.

The graphical interface of EPRSIM-WIZ (presented in Figure 2) adapts to different arrangements of experimental and simulation data in such a way that any the software can be upgraded for new models or new type of complex experiments easily.
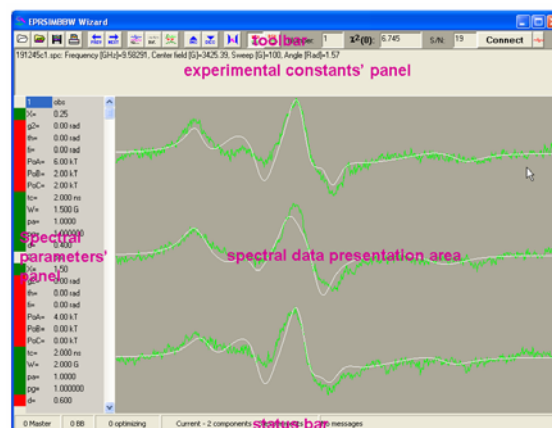

Figure 2: EPRSIM-WIZ user interface with indicated different parts

EPRSIM-BBW

The purpose of EPRSIM-BBW is primarily to perform automatic optimization of series of spectra. Multithreading principle and calculation in background are the two most important properties of this program.

EPRSIM-BBW can be run for optimization of problems defined in a single MTP-file. However, EPRSIM-BBW can be run also in "stay-resident" mode in which it is continuously checking the specified directory for any new tasks to be performed. The EPRSIM-BBW modes are triggered via various switches in command prompt. For the detailed description of switches see implementation notes and manuals in the software package. Here are the two examples that should illustrate the usage of EPRSIM-BBW:

1. To run EPRSIM BBW in "single MTP optimization" mode, pointing to the »c:\temp\test.mtp« file containing problems for optimization, allowing at maximum 2 parallel threads and forcing 20 independent runs on each problem defined in MTP file, one should input the following line:

   eprsimbbw.exe -p=c:\temp\test.mtp -n=2 -k=20

   When the optimization is finished, the optimized results are written in terms of "test1.mtp" file containing best fits and POP subdirectory containing population files for GHOST construction, and EPRSIM-BBW program terminates.

2. To run EPRSIM BBW in "stay-resident" mode, pointing to the »c:\temp« directory that eventually contains MTP files with problems for optimization, allowing at maximum 2 parallel threads and forcing 20 independent runs on each problem defined in MTP file, one should input the following line:

   eprsimbbw.exe -d=c:\temp -n=2 -k=20

   After EPRSIM-BBW is started, it checks the »c:\temp« directory every 30 second for new MTP files to start optimization. When MTP file (together with appropriate spectral data files) is copied to »c:\temp« directory, EPRSIM-BBW will automatically start optimization at lately in 10 seconds. When finished, it starts to scan the directory again.

   EPRSIM-BBW can be started directly from command line (prompt) or from Windows Schedule to enable pure background running. EPRSIM BBW can be terminated through Task Manager or Windows Schedule (if used).

   File transfer to the directory can be from anywhere (with appropriate permissions) through the local network. When several new MTP files are found in the directory, the oldest one starts first.

EPR-Convert

The purpose of EPR-Convert is to convert Bruker binary spectral files into ASCII files in order to be used in EPRSIM-C programs or plotted with any other program.

EPR-Convert can convert Bruker SPC (OS-9 file format) and DTA (Elexsys file format). Note that in case of DTA files, 2D experimental data are converting by splitting data in several spectral files, denotes by "_slice#" with # as a slice index.

EPR-Convert can also convert several spectral files into a single multi-column ASCII file. To distinguish between the data, the columns are indexed by the appropriate filenames.

GHOSTMAKER

The purpose of GHOSTMAKER is to condense the multi-run dHEO solutions into GHOSTs and representative / distinguishable groups of solutions.

In general, one can extract at least the best solution from each run when genetic algorithm is applied for optimization. As this kind of optimization is purely stochastic, one must compare at least the best solutions from each run to determine the uncertainty of solutions. If the proposed model complexity is large enough the uncertainty represent pure deviation of solutions. However, if proposed complexity is too low and the optimization finds larger regions of phase space to describe the spectral data, these uncertainties represent also the distribution of solutions.

Basic HEO algorithm always strongly converges to one good region because of the strong effect of the crossover operator. Therefore in the case when proposed model complexity is too low, one should never extract more than only best-fit per run. However, with dHEO, with shaking operator introduced to prevent crowding, one can extract more different solutions than just best-fit per run. In that way, the reduction in computational time for EPRSIM-BBW is drastic.

Despite the way, how the solutions are generated, either by HEO or dHEO, large number of solutions (like 200 per problem) has to be condensed taking into account different efficiency of individual runs as well as probability with which the solution regions are found.

The χ^2-filtering and density filtering are the two main properties of GHOST condensation. In local solution density calculation as well as in the **grouping of solutions**, the principle of solutions' neighborhood is implemented. For the detailed description see implementation notes and manual for EPRSIM-C software package. In the final step of GHOST condensation the groups of solutions have to be quantified for their proportion. First approximation is made according to the values condensed out from individual solutions. However, the final determination of proportions is done according to the spectral weights found by additional run of EPRSIM-BBW after sending up to 5 descriptive solutions of GHOST condensate as the spectral components to EPRSIM-BBW.
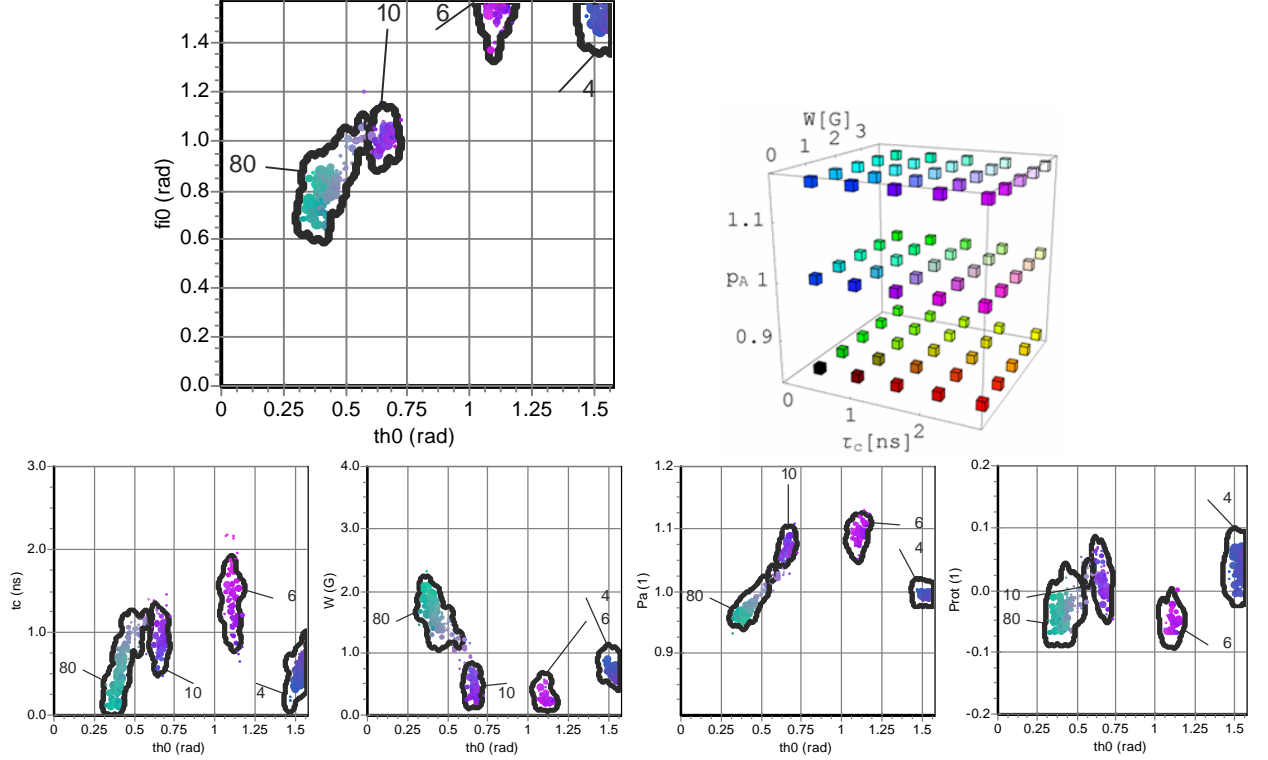
Figure 3: Typical main GHOST for MES model, color legend, and additional GHOST cross-sections; note that tc, W and pA cross-sections are hidden in the main GHOST by color coding

After GHOST condensation the results are **represented by GHOSTs** (GHOST diagrams, see figure 3). Currently, GHOST condensation is working for MEM, MES and LLE simulation models for which the x-axes parameter is chosen to be order parameter S_z, main cone angle ϑ, and spin-exchange rate W_{exc}, respectively for the three models. The three of the other parameters: rotational correlation time τ_c, additional broadening W, and polarity correction p_A, are always coded by RGB color directive, i.e. with red, green and blue color intensity coding the τ_c, W, and p_A, respectively. Additional parameters, like φ, prot, etc. are represented on the y-axes.

It should be stressed that when using GHOST condensation methodology, certain rules has to be obeyed, e.g. taking into account signal-to-noise ratio, checking uniformness of run-contribution histogram, adopting the minimal threshold density, etc. Their detailed description could be found in the implementation notes and manual of the EPRSIM-C software package.

References

- STRANCAR, Janez, SENTJURC, Marjeta, SCHARA, Milan Valter. Fast and accurate characterization of biological membranes by EPR spectra. *J. magn. reson. (San Diego, Calif., 1997 : Print)*, 2000, vol. 142, 254-265.

- FILIPIC, Bogdan, STRANCAR, Janez. Tuning EPR spectral parameters with a genetic algorithm. *Applied soft computing*, 2001, vol. 1, no. 1, 83-90.
- STRANCAR, Janez, SCHARA, Milan Valter, PECAR, Slavko. New EPR method for cellular surface characterization. *J Membr Biol*, 2003, vol. 193, 15-22.
- STRANCAR, Janez, KOKLIC, Tilen, ARSOV, Zoran. Soft picture of lateral heterogeneity in biomembranes. *J Membr Biol*, 2003, vol. 196, 135-146.
- FILIPIC, Bogdan, STRANCAR, Janez. Evolutionary computational support for the characterization of biological systems. in: FOGEL, Gary B. (ed.), CORNE, David (ed.). *Evolutionary computation in bioinformatics*. Amsterdam [etc.]: Morgan Kaufmann; Oxford: Elsevier Science, 2003, 279-294.
- STRANCAR, Janez, KOKLIC, Tilen, ARSOV, Zoran, FILIPIČ, Bogdan, STOPAR, David, HEMMINGA, Marcus A. Spin label EPR-based characterization of biosystem complexity. *J. chem. inf. comput. sci.*, 2005, vol. 45, 394-406.
- KAVALENKA, Aleh A., FILIPIC, Bogdan, HEMMINGA, Marcus A., STRANCAR, Janez. Speeding up a genetic algorithm for EPR-based spin label characterization of biosystem complexity. *J. chem. inf. comput. sci.*, 2005, vol. 45, 1628-2635.
- STOPAR, David, STRANCAR, Janez, SPRUIJT, Ruud B., HEMMINGA, Marcus A. Exploring the local conformational space of a membrane protein by site-directed spin labeling. *J. chem. inf. comput. sci.*, 2005, vol. 45, 1621-1627.
- EPRSIM-C WEB page: http://www.ijs.si/ijs/dept/epr/EPRSIMC_overview.htm