Neural Network Simulation of Energy Transfer Processes in a Membrane Protein System

Petr V. Nazarov[1], Vladimir V. Apanasovich[1], Vladimir M. Lutkovski[1], Marcus A. Hemminga[2], Rob B. M. Koehorst[2]

[1] Department of Systems Analysis, Faculty of Radiophysics, Belarusian State University, Skaryna Ave 4, Minsk 220050, Belarus. E-mail: lutkovski@bsu.by
[2] Laboratory of Biophysics, Wageningen University, Dreijenlaan 3, 6703 HA, Wageningen, The Netherlands. Website: http://ntmf.mf.wau.nl/hemminga/

Abstract. The method of the direct application of an artificial neural network for modelling of a complex system is developed with the purpose of speeding up the optimisation procedure for determination of system parameters. The method provides a significant decrease in simulation time. Moreover the artificial neural network produces a smooth approximation of stochastic simulation results and consequently it reduces the level of stochastic errors. The developed algorithm is applied to model the fluorescence resonance energy transfer within a system of M13 major coat protein mutants embedded in a membrane.

1. Introduction

Computer modelling has become a standard method for the parameter determination and study of complex systems and processes. Especially in systems that cannot be described by a set of analytical equations Monte Carlo simulation provides a general approach to such a problem. However, computer simulation has several weak points. One of them is the high time cost of its realisation, which makes it difficult to use an optimisation technique for the determination of model parameters. At the same time, it is known that it is possible to approximate any existing smooth function using an artificial neural network (ANN) with a sufficient number of layers and neurons [3, 5, 1].

In this paper we propose to apply an ANN to reproduce a simulation model. This makes it possible to use the ANN to fit the model parameters to an experimental data set, and to strongly speed up the fitting procedure. Here we demonstrate the application of this idea to the simulation model of energy transfer processes between fluorophores in a membrane protein system.

2. Theory

Let us consider the following problem. A system Θ has a set of input parameters P, and a set of output values F. Some input parameters are known (P_0) and others

should be defined (P_X), so that P=[P_0, P_X]. In this case, the system described performs the following operation:

$$\Theta(P_0, P_X) = F \,. \tag{1}$$

Usually the determination of the unknown parameters P_X is carried out in the following way:

1. A set of F is obtained experimentally.
2. A model Ξ of system Θ is created. It performs the operation:

$$\Xi(P_0, P_X) = F^* \tag{2}$$

3. An initial estimation is made for P_X.
4. An optimisation algorithm, using a variation of parameters P_X, is executed to minimize the error $||F^*-F||$.

This very general algorithm is represented in Fig.1.

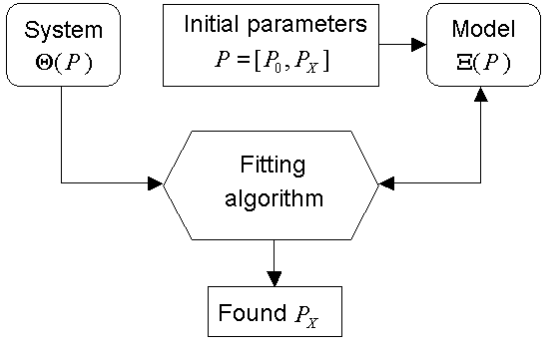

Fig.1. Methodology of parameter determination

The most crucial problem of this scheme is the time-consuming operation (2). In some cases this approach is not useful at all, because the time of optimisation becomes non-realistic. To significantly speed up the algorithm, we propose the application of an ANN that replaces model (2).

Continuous functions can be uniformly well approximated by linear combinations of sigmoidal functions, as was independently shown by Cybenko [3], and Hornik, Stinchcombe, and White [5]. The error in the approximation of functions by an ANN is bounded [1]. For our case this means that it is possible to replace the operation (2) by a neural network transform

$$\Psi(P_0, P_X) = F^* \,. \tag{3}$$

The computation time needed for transform (3) is much less than for operation (2). The suggested neural network approach to modelling is illustrated in Fig. 2. In this scheme the generation of a representative training set is the most time-consuming procedure. To be sure the training set is representative we used the following algorithm.

Two subsets of training pairs are taken: a deterministic and a stochastic one.

Step 1. The generation of a "boundary" parameters set. For each parameter 3 values are taken: minimal, average and maximal. Then all possible combinations of these values are generated which provides the first training subset. For example, for a system with 4 input parameters this gives 3^4=81 training pairs in this subset.
Step 2. The generation of a primary training subset. In this step input parameters are randomly selected within a certain range. In our numerical experiments we take up to 5000 of training pairs.
Step 3. The determination of similar training pairs. Due to the random selection of parameters, some training vectors can be close to each other. Their use for training does not lead to an increase of precision, only to an increase of training time.

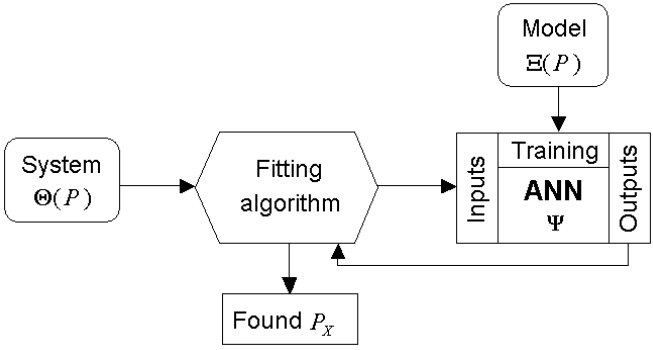

Fig.2. The direct application of an ANN to modelling

The method described is used for the simulation of dipole-dipole energy transfer processes between membrane proteins.

3. Model Description

The fluorescence resonance energy transfer (FRET) method has been applied to several problems in biology as a means of estimating intra- and intermolecular distances in macromolecular systems. Especially it has been used for the determination of the conformation and association of proteins embedded into a phospholipid bilayer. [6] The idea of FRET spectroscopy consists of labelling macromolecules with fluorescent probes of two types: donors with a relatively long lifetime of the excited state and acceptors with a short lifetime. The emission spectrum of the donor and the absorption spectrum of the acceptor should overlap. The donors are excited by a light source and some of them transfer excitation energy to the acceptors. From the resulting emission spectra, it is possible to define an efficiency of the energy transfer processes using Föster's theory [4]. The experimental system consists of mixed lipid vesicles of *dioleoyl phosphatidylcholine* (DOPC) and *dioleoyl phosphatidylglycerol* (DOPG) with inserted major coat proteins of the M13 bacteriophage [7]. In this system there are two types of proteins – wild, containing only one *tryptophane* amino acid (donor) and mutant proteins, containing both *tryptophane* and *AEDANS* (acceptor).

Because of its complexity in these reconstituted protein-lipid systems, the problem of determining the association and conformation of the M13 coat protein cannot be solved analytically. All analytical solutions that are known today have very serious simplifications and limitations. Therefore we used a simulation model. In this model a biological membrane is approximated by a two-dimensional periodic structure with a hexagonal packing of the lipids in which proteins are randomly distributed (Fig. 3 **a**). The area occupied by each protein on the membrane surface is assumed to be equal to that of a lipid. The distance between two nearest molecules on the grid is 5.6 Å and the thickness of the lipid bilayer is 30 Å. M13 coat protein mutants are approximated by hexagonally situated rods with a constant location of the donor (D) and variable location of the acceptor (A) (Fig. 3 **b**).

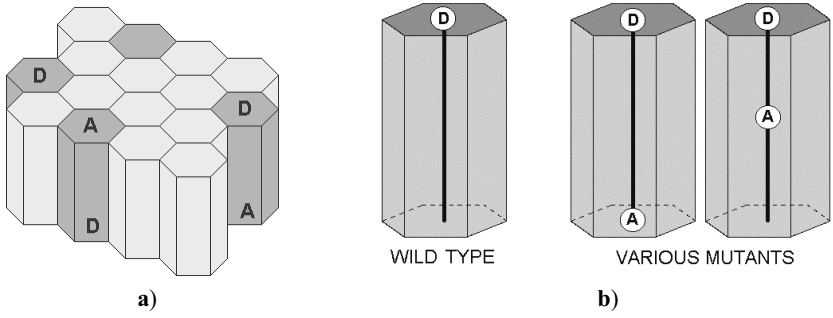

Fig 3. Model of a membrane (**a**) and membrane proteins (**b**) with fluorescent labels

The input parameters of the model are:

1) Concentrations of wild-type proteins and mutants.
2) Location of the acceptor in the protein.
3) Coefficient of protein-protein association (the probability that a selected protein is located in immediate proximity to another one).
4) Föster distance (the photophysical parameter, which describes the intensity of energy transfer [4]).

The intensity of fluorescence and energy transfer efficiency are taken as output values. Because of the simulation nature of the model, the resulting output contains stochastic errors. Therefore simulations are run several times to reduce these errors. This results in an increase of calculation time. In our numerical experiments one complete calculation takes up to 2 min.

4. Results

A three-layer perceptron is used to reproduce the described simulation model. Some statistic experiments are conducted to define the optimal number of neurons in hidden layers. This number depends on the number of variable input

parameters. So, for 3 variable parameters the ANN should have 10-12 neurons in hidden layers, for 4 parameters 14 and for 5 parameters – 16. These values were experimentally obtained. In our case the application of a larger number of neurons does not lead to a decrease of the error for a test set.

The ANN is trained by the Back Propagation Error algorithm with the Levenberg-Marquardt optimisation technique [2]. To avoid overtraining after each 10 epochs, the ANN is tested on a set of control input-output pairs (200 elements). If the results of testing did not improve for a certain time, the training procedure is stopped. All calculations were made in MATLAB® 6.1 with the Neural Networks Toolbox on a PC with Intel Pentium III-850 CPU. The time costs of the ANN method application are shown in Table 1.

Table 1. The time costs of the ANN method

Task	3 variable parameters	4 variable parameters
Training subset 1 generation	30 min	90 min
Training subset 2 generation	12 hr	22 hr
Training	6 min	12 min
ANN simulation	$5 \cdot 10^{-4}$ s	$8 \cdot 10^{-4}$ s

The average time of the calculation conducted by the simulation model is 48 sec. To obtain the same result by the ANN calculation it takes about $6 \cdot 10^{-4}$ s, depending on the number of neurons in the hidden layers. This means that the calculation is speeded up by approximately a factor of 10^5.

After execution of the training procedure the mean relative square error on the training set varied from 0.8 to 1.2% depending on the task. The resulting errors on the test input-output set varied from 0.9 to 1.5%. To prove the quality of the method used, comparison were made using the simulation model and the ANN. One result is represented in Fig 4, which shows an excellent agreement.

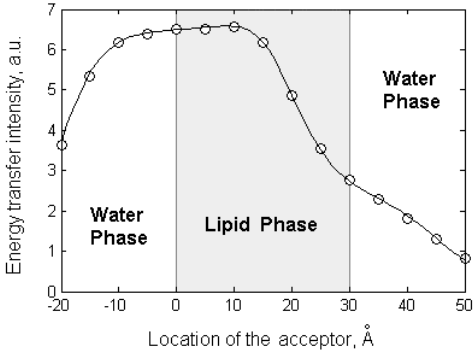

Fig. 4. The dependence of the energy transfer intensity on the location of the acceptor label (assuming no protein aggregation). The circles show the computer-simulated values, and the line shows the neural network approximation

Fig. 4 shows that an ANN is able to build a smooth interpolation of the dependences of fluorescence intensity from input parameters of the model. Due to high speed of the ANN calculation it is possible to obtain smooth dependences. These dependences are quite complex and so they hardly can be described by a single analytical function with physically meaningful parameters. Moreover an ANN approximation of a model allows calculating numerical values of derivatives and hence it allows using a gradient fitting technique.

5. Conclusions

The use of a trained ANN in the modelling presented here results in a gain in computing time by a factor of 10^5. This allows using optimisation technique to determine parameters of the system considered. Moreover the ANN produces a smooth approximation of the results of a stochastic simulation. Thus it decreases the level of stochastic errors. Due to this smooth dependency it is possible to use a standard optimisation technique, such as gradient search, for parameter determination.

The imperfections of described algorithm are the following. It works only when the number of variable parameters is quite small (in our experiments – up to 5). Furthermore ANN approximation of a system may contain some inaccuracies.

Nevertheless, the method of ANN modelling can be used as a very good tool for determination of initial estimations of parameters. It is not limited to biophysical problems, and can be applied to any situation where the scheme from Fig.1 can be used.

References

1. Barron AR (1993) Universal approximation bounds for superpositions of a sigmoidal function. IEEE Trans on Inform Theory, 39:3:930-945
2. Bishop M (1997) Neural Networks for Pattern Recognition. Clarendon Press, Oxford
3. Cybenko G (1989) Approximations by superpositions of a sigmoidal function. Math Contr Signals Syst, 2:304-314
4. Föster T (1948) Intermolecular energy migration and fluorescence. Ann Phys 2:55-75
5. Hornik K, Stinchcombe M, and White H (1989) Multilayer feedforward networks are universal approximators. Neural Networks, 2:359-366
6. Koppel D, et al. (1979) Intramembrane Position of Membrane-Bound Chromophores Determined by Excitation Energy Transfer. Biochemistry 18:5450-5457
7. Spruijt R, et al. (2000) Localization and rearrangement modulation of the N-terminal arm of the membrane-bound major coat protein of bacteriophage M13. Biochim Biophys Acta 1509:311-323