

Scientific paper

Structure Elucidation of Transmembrane Proteins using Public-Available Databases and Experimental Data on Competitive Inhibition

Amrita Roy Choudhury,¹ Špela Župerl,¹ Sabina Passamonti²
and Marjana Novič^{1,*}

¹ Laboratory of Chemometrics, National Institute of Chemistry, Hajdrihova 19,
SI-1000 Ljubljana, Slovenia

² Department of Life Sciences, University of Trieste, Via L. Giorgieri 1, I-34127 Trieste, Italy.

* Corresponding author: E-mail: marjana.novic@ki.si

Received: 09-03-2011

Dedicated to Professor Dušan Hadži on the occasion of his 90th birthday

Abstract

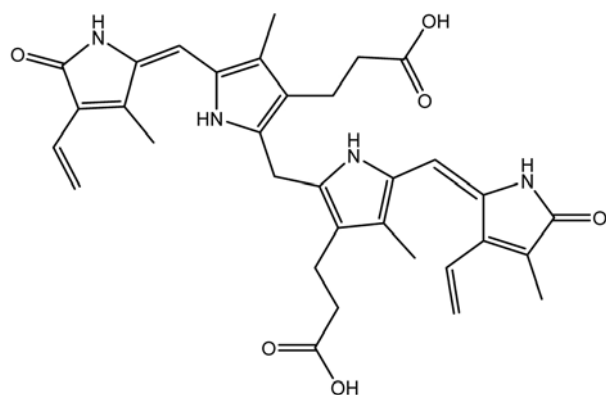
We present an approach towards structure elucidation of biliranslocase, the membrane protein which transports bilirubin from blood to liver cells. The sequence and secondary structure information of transmembrane segments of proteins with known 3D structure is exploited to predict the transmembrane domains of structurally unresolved target protein. With the help of known structures the transmembrane domains are encoded in such a way that it is possible to group and classify them with respect to their specific sub-structural characteristics and to build a model for prediction of transmembrane segments. We have shown that the model for prediction of transmembrane segments proposed four transmembrane alpha helices, each containing around 20 amino acids. This result is partially confirmed with experimental studies using particular antibodies corresponding to parts of amino acid sequences of biliranslocase. In order to shed light on the biliranslocase transport mechanism, we also tested a set of non-congeneric compounds for their competitive inhibition constants in the investigated protein-substrate system. The information about chemical structure of small molecules that either pass or block the transmembrane path enabled by biliranslocase helps us to build a hypothesis about the transport mechanism of the studied biological system.

Keywords: Transmembrane protein, biliranslocase, neural network, predictive modeling, inhibition constant

1. Introduction

Biological membranes form multiple barriers through which both drugs and toxic molecules enter the organism. Despite the difficulties encountered in biomembrane research, investigating mechanisms of membrane transport is basic to understanding the bioactivity of most drugs. Within this scope, the information about a 3D structure of a membrane transporter is of great importance in tackling the study of the protein and small molecule transport mechanism. Only a limited number of membrane transport proteins have been already solved experimentally for their 3D structure. However, it is not trivial to obtain the X-ray structure of a membrane protein due to a

complex and demanding procedure of heterologous production, purification and crystallization, which leaves the tertiary structure of a large number of membrane proteins still unresolved. Transmembrane protein molecules are difficult to crystallize due to their amphiphilic characteristics – hydrophobic transmembrane segments and hydrophilic loops. In the absence of the experimentally obtained 3D structures, in-silico methods may fill the information gap and offer a possibility to hypothesize the transport mechanism, providing the in-silico modelling is based on experimental evidence about the transport potency for structurally diverse molecules. In turn, for the structural characterization of membrane proteins, modeling of their transport activity for molecules of diverse chemical structure may offer a great contribution.



Scheme 1. Chemical structure of bilirubin.

The aim of this work was to investigate the mechanism of transport and structural details of bilitranslocase, the transporter of organic anions, such as bilirubin (Scheme 1), from blood to liver cells. It is inhibited competitively by number of structurally diverse molecules, such as anthocyanins and their mono- and di-glycosylated derivatives. Bilitranslocase may thus be involved in the bioactivity of flavonoids; it also makes it a good candidate for being a transport target for other polar molecules with therapeutic application.^{1,2} One of the goals of this work is to enrich the experimental data on the transport activity of bilitranslocase for structurally diverse compounds and develop a data-driven model for prediction of inhibition constants of structurally modified molecules. The benefit of the computational model is not only the prediction of inhibition constants for new molecules; the interpretation of structural descriptors and site-specific variables chosen in the procedure during the model optimization help us to better understand the influences on the transport mechanism. The lack of knowledge of the secondary structure of bilitranslocase makes any explanation of the transport mechanism challenging. For this reason the next goal is to build a model of 3D structure of bilitranslocase. The final aim is to combine two approaches: first, the prediction of transmembrane segments of the protein based on the mathematical descriptors obtained from the information of membrane proteins of known 3D structure available in public databases (Protein Data Bank of Transmembrane Proteins), and second, the information about the transport mechanism from experimentally tested set of small molecules for their competitive inhibition of bilitranslocase.

2. Data Compilation

Sequences and information regarding transmembrane regions of integral-membrane proteins with known 3D structures were collected from public databases PDB and PDBTM.^{3,4} We considered only alpha transmembrane proteins, since alpha helix is the more common secondary

structural feature of these proteins. The initial dataset consisted of 824 proteins (all such proteins in PDBTM database as of Jan. 23, 2009). In the first step, sequences of proteins with low-resolution and theoretically determined structures were removed. Identical sequences of multimeric and same proteins were omitted as well to avoid data redundancy. We then separated the transmembrane and non-transmembrane regions of each of the protein chains in the reduced dataset. Lastly, non-transmembrane regions were further divided into shorter segments of length comparable to the length of transmembrane regions, as the model gives optimal results if it is built on objects with comparable parameters. The final dataset consisted of 5800 labeled protein segments, 2545 transmembrane and 3255 non-transmembrane. To compile and refine the dataset, to prepare the counter-propagation neural network (CP-ANN)^{5,6} input, and analyze the prediction model output we developed codes in Perl5.10.0 programming language.⁷

A set of non-congeneric compounds was tested in order to obtain bilitranslocase inhibition constants (K_i). In Table 1 55 compounds (nucleobases, nucleosides, nucleotides (ID numbers from 1 to 41), and various endogenous compounds and drugs (ID numbers from 42 to 55)⁸ are listed together with their ID numbers, molecular weight, type of activity, *i.e.* inactive (I), competitive (C) and non-competitive (NC) inhibition, and experimental values of inhibition constants (K_i) with their standard errors (*SE*). Experimental data obtained by testing the interaction of bilitranslocase with anthocyanines (ID numbers from 101 to 122) and flavonols (ID numbers from 123–143), together 43 compounds,² are given in Table 2. The interaction was assessed³ by evaluating the kinetics of inhibition of bilitranslocase transport activity. The experiments for the determination of bilitranslocase inhibition constants were performed with a series of substrate (sulphobromophthalein) concentrations, while the investigated molecules were added in stoichiometric concentrations. The experimental results were the basis for a data-driven modelling study using artificial neural networks.

Table 1. List of 55 compounds, nucleobases, nucleotides, and nucleosides, with their ID numbers (ID = 1–55), type of activity, and experimental inhibition constants with their standard errors (*SE*).

ID	Compound	Activity	$K_i \pm SE$ (mmol/L)
1 ¹	Adenine	I	–
2 ¹	Adenosine	I	–
3 ¹	Adenosine 3'-monophosphate	C	0.95±0.07
4 ³	Adenosine 5'-monophosphate	C	2.63±0.19
5 ¹	Adenosine 3', 5'-cyclic monophosphate	I	–
6 ²	Adenosine 5'-diphosphate	C	1.42±0.10
7 ¹	Adenosine 5'-triphosphate	C	0.385±0.03
8 ²	Adenosine-5'-diphosphoglucose	I	–

ID	Compound	Activity	$K_i \pm SE$ (mmol/L)
9 ¹	Adenosine 5'-(α,β -methylene) diphosphate	C	1.31±0.14
10 ¹	Adenine 9- β -D-arabinofuranoside	NC	3.76±0.38
11 ¹	Adenosine 3'-phosphate 5'-phosphosulfate	C	0.148±0.01
12 ²	S-(5'-Adenosyl)-L-homocysteine	C	0.44±0.02
13 ¹	S-(5'-Adenosyl)-L-methionine	C	0.408±0.03
14 ³	Guanine [#]	Not soluble	/
15 ¹	Guanosine	I	–
16 ¹	Guanosine 5'-monophosphate	C	13.92±1.38
17 ¹	Guanosine 3', 5'-cyclic monophosphate	I	–
18 ¹	Guanosine 5'-diphosphate	C	4.55±0.34
19 ¹	Guanosine 5'-triphosphate	NC	7.66±0.09
20 ¹	Uracil	I	–
21 ¹	Uridine	C	2.58±0.24
22 ¹	Uridine 5'-monophosphate	C	4.13±0.38
23 ²	Uridine 5'-diphosphate	C	3.10±0.37
24 ¹	Uridine 5'-triphosphate	C	1.425±0.08
25 ¹	Uridine 5'-diphosphoglucose	I	–
26 ¹	Uridine 5'-diphosphogalactose	C	2.47±0.27
27 ³	Uridine 5'-diphosphoglucuronic acid	I	–
28 ²	Thymine	I	–
29 ¹	Thymidine	I	–
30 ²	Thymidine 5'-monophosphate	C	3.71±0.18
31 ¹	Thymidine 5'-diphosphate	C	2.23±0.15
32 ¹	Thymidine 5'-triphosphate	C	1.45±0.08
33 ¹	Cytosine	I	–
34 ¹	Cytidine	I	–
35 ²	Cytidine 2'-monophosphate	I	–
36 ³	Cytidine 2':3'-cyclic monophosphate [#]	/	/
37 ¹	Cytidine 5'-monophosphate	I	–
38 ³	Cytidine 5'-diphosphate [#]	/	/
39 ³	Cytidine 5'-triphosphate [#]	/	/
40 ³	Cytosine β -D-arabinofuranoside [#]	/	/
41 ³	Cytosine β -D-arabinofuranoside 5' monophosphate [#]	/	/
42 ¹	Uric acid	C	1.50±0.18
43 ¹	Ouabain	I	–
44 ¹	Aucubin	I	–
45 ²	Loganin	I	–
46 ³	Verbenalin	I	–
47 ¹	Isovitexin	I	–
48 ¹	Vitexin-2'-O-rhamnoside	I	–
49 ¹	Cibacron Blue F3G-A	C	0.00347±0.00032
50 ²	Digoxin	I	–
51 ¹	Taurocholate	I	–
52 ¹	Sulfobromophthalein	C	0.00532±0.00063
53 ³	Thymol Blue	C	/
54 ¹	Bilirubin	C	0.00111±0.00001
55 ²	Biliverdin	C	0.00111±0.00002

[#] compound was not tested I – inactive compound

C – competitive inhibitor NC – non-competitive inhibitor

¹ training set ² test set ³ external validation set

Table 2. List of 43 compounds, anthocyanines and flavonols, with their ID numbers (ID = 101–143), type of activity, and experimental inhibition constants with their standard errors (SE).

ID	Compound	Activity	$K_i \pm SE$ (mmol/L)
101	Pelargonidin	C	22.21±1.65
102	Cyanidin	C	17.55±1.68
103	Delfinidin	C	5.27±0.38
104	Peonidin	C	6.23±0.51
105	Petunidin	C	7.57±0.99
106	Malvidin	C	7.20±0.40
107	Pelargonidin 3-O-D-glucopyranoside	C	2.79±0.18
108	Cyanidin 3-O-D-glucopyranoside	C	5.78±0.39
109	Delfinidin 3-O-D-glucopyranoside	C	8.57±0.20
110	Peonidin 3-O-D-glucopyranoside	C	1.83±0.19
111	Petunidin 3-O-D-glucopyranoside	C	4.03±0.19
112	Malvidin 3-O-D-glucopyranoside	C	1.42±0.13
113	Pelargonidin 3,5-di-O-D-glucopyranoside	C	6.42±0.29
114	Cyanidin 3,5-di-O-D-glucopyranoside	C	5.77±0.39
115	Peonidin 3,5-di-O-D-glucopyranoside	C	6.81±0.77
116	Malvidin 3,5-di-O-D-glucopyranoside	C	6.36±0.45
117	Cyanidin 3-O-L-arabinopyranoside	C	9.16±0.99
118	Cyanidin 3-O-D-galactopyranoside	NC	35.22±0.58
119	Malvidin 3-O-(6-O-acetyl)-D-glucopyranoside	NC	58.33±0.09
120	Delfinidin 3,5-di-O-D-glucopyranoside [#]	/	/
121	Petunidin 3,5-di-O-D-glucopyranoside [#]	/	/
122	Malvidin 3-O-(6-O-p-coumaroyl)-D-glucopyranoside	I	–
123	Galangin	NC	60.6±1.0
124	Kaempferol	NC	63.9±3.4
125	Quercetin	NC	131.6±3.8
126	Myricetin	C	79.6±3.6
127	Syringetin	C	21.1±1.7
128	Rhamnetin	I	–
129	Isorhamnetin [#]	/	/
130	Quercetin 4'-glucopyranoside	I	–
131	Quercetin 3,4'-diglucopyranoside	I	–
132	Quercetin 3-glucopyranoside	I	–
133	Quercetin 3-xyloside	I	–
134	Quercetin 3-rhamnoside	I	–
135	Quercetin 3-galactoside	I	–
136	Quercetin 3-O-glucopyranosyl-6"-acetate	I	–
137	Quercetin 3-O-sulfate	I	–
138	Isorhamnetin 3-glucoside	I	–
139	Isorhamnetin 3-O-rutinoside	I	–
140	Kaempferol-glucoside	I	–
141	Kaempferol 3-O-rutinoside	I	–
142	Syringetin 3-galactoside	I	–
143	Syringetin 3-glucoside	I	–

I – inactive compound C – competitive inhibitor

NC – non-competitive inhibitor

3. Results and Discussion

Part I: Prediction of Transmembrane Segments

We prepared the amino acid (AA) adjacency matrix of each transmembrane sequence. The 20-element row-sum vector of AA adjacency matrix was taken for characterization of protein sequence (protein descriptors). Our choice was based on assumption that as protein structure and function depend on its sequence, the transmembrane proteins must have some specific sequence patterns that give them characteristic folds and properties, distinguishing them from globular proteins. The sequence patterns also inherently represent patterns in which amino acids of particular hydrophobicity occur in the sequence.

The error of prediction of transmembrane segments for the external validation set was below 10%. Moreover, as the comparison shows, the model is able to predict all the proposed transmembrane regions for an unknown protein with far better accuracy than most of the other available methods.⁹ In Table 3 an example of testing seven transmembrane and seven globular proteins for potential presence of alpha transmembrane regions is given.

Table 3. Seven transmembrane and seven globular proteins from PDB database tested for presence of transmembrane (TM) alpha helices.

Alpha transmembrane proteins					Globular proteins		
PDB Id	Total TM	Pred TM	False positive	False Negative	PDB Id	Helices present	Predicted as TM helices
<i>2npk</i>	11	7	0	4	<i>3gak</i>	14	0
<i>1bha</i>	2	2	0	0	<i>3h9e</i>	13	1
<i>BTL</i>	–	4	–	–	<i>3b97</i>	21	0
<i>lotu</i>	10	6	1	4	<i>3cls</i>	10	0
<i>2bhw</i>	3	3	0	0	<i>3h1v</i>	19	0
<i>2ahy</i>	2	2	0	0	<i>2wu8</i>	31	1
<i>3c9m</i>	7	7	0	0	<i>li7y</i>	9	0

As shown in Table 3, of the six transmembrane proteins of known transmembrane regions, 27 out of 35 regions are predicted correctly. There are only eight false negative and one false positive. Moreover, all the four probable transmembrane regions of bilitranslocase are predicted correctly. In case of globular proteins, only two false positive predictions were obtained, what means that two alpha helices out of 117 present in seven proteins were predicted as transmembrane segments. The error in these seven cases was thus 1.7%, however, for testing a larger number of proteins some improvement of the developed software is needed towards automation of the procedure.

The model for prediction of alpha transmembrane segments was challenged with bilitranslocase containing 340 amino acid residues. The model proposed four transmembrane alpha helices, each containing around 20 amino acids. After additional constraints obtained from a detailed statistical analysis of position-specific amino acid

abundances¹⁰ the following transmembrane alpha helices in bilitranslocase were predicted:

TM1	24–48	FTKCILVSSSFLLFYTLPHGLLED
TM2	75–94	FCLFVATLQSPFSAGVSGLC
TM3	220–238	GSVQCAGLISLPIAIEFTY
TM4	254–276	PNIFPLIACILLLSMNSTLSLFS

The remaining protein consists of five loops given below:

LOOP	1–23	MLIHNWILTFISIFREHPSTVFQI
LOOP	49–74	LMRRVGDSDLVDLIVICEDSQGQHLSS
LOOP	95–219	KAILLPSKQIHVMIQSVDSLIGITNS LTNEQLCGFGFFLN VKTNLHCSRIPLITNLFLSARHMSL DLENSVGSYHPRMIW SVTWQWSNQVPAFGETSLGFGM FQEKGRHQNYEFPCCRCIGTCGR
LOOP	239–253	QLTSSPTCIVRPWRF
LOOP	277–340	FSGGRSGYVLMMLSSKYQDSFTS KTRNKRENSIFFLGLNTF TDFRHTINGPISPLMRSLTRSTVE

As the second TM (TM2) region is just after the bilirubin binding motif (residue 65–75, as evident from antibody studies),^{11,12} it indicates that TM2 may form the

wall of the transporting channel with the binding motif close to it that initiates the active transport.

Part II: On Transport Mechanism of Small Molecules

The information about 3D chemical structures (minimal energy conformation from Mopac)¹³ of the molecules tested for their bilitranslocase inhibition constants were represented by molecular descriptors calculated with the Codessa software.¹⁴ The molecular descriptors,¹⁵ together with the corresponding bilitranslocase inhibition constants were used to train the counter-propagation neural network, designed for classification and prediction purposes. The Kohonen neural network was applied for visualization, clustering purpose and for the division of the data set.^{5,6} From distribution of all 98 compounds in the Kohonen top map (see Figure 1) we have observed that the nucleobases, nucleotides and nucleosides are

completely separated from the flavonoids (anthocyanins and flavonols). For that reason we have divided dataset into two datasets; nucleobases and their derivatives (IDs from 1 to 55) and flavonoids (IDs from 101 to 143). The two datasets were modeled separately but with the same modeling procedure.

case with the pK_i equal to $5.32 \pm 0.63 \mu\text{M}$. The 3D structure was prepared in the same way as for all other molecules in the study. Once structural descriptors were obtained, 155 out of 353 descriptors were retained in the model building for further variable reduction. The nature of the counter-propagation neural network allows building mo-

50 51 53	43	111 112	108 109 118	117	104	101 102 103	123 124 125 126		1 20 28 33
54 55		119	107 110		105 106				14
113 114 115 116 120 121	122					127 128 129			42
48 131 139			44 45 46					137	
141					21 29				30
142 143	136 138	132 133 134 135 140		34 40			35 37 41	22	31
52					15	36		38	23
47 130		12		2 10		17	16		24 32 39
					5		18	19	
49		13	9		3 4	6	7 11	8	25 26 27

Figure 1. Distribution of 98 objects in the Kohonen top map with dimension 10×10 . ID numbers from (1 to 41) in black, nucleobases, nucleotides and nucleosides; ID numbers (from 42 to 55) in blue, various endogenous compounds and drugs; ID numbers (from 101 to 122) in green, anthocyanins; ID numbers (from 123–143) in red, flavonols.

The Kohonen neural networks were applied for the division of the two datasets into training (TR), test (TE) and validation (V) sets.^{16,17} For predictive purposes the Counter-propagation neural network (CP-ANN) was used. Optimized models were validated by previously determined validation set. Genetic algorithm (GA)¹⁸ for variable selection was introduced and detailed investigation of selected descriptors was performed. The root mean squared error of prediction (RMSEP) of the negative logarithm of inhibition constants (pK_i) of flavonoids was equal to 2.2. For a small dataset such as that investigated in the flavonoids study this error is expected. We are especially satisfied with the results after inspection of the obtained clusters of training compounds in the Kohonen map: the tested molecules were correctly placed in structurally similar clusters. That result proves that the set of descriptors is properly chosen for prediction of K_i value. The final validation of the neural network model [2] was performed by testing the inhibition activity of bromosulphothalein, which is an established substrate of bilitranslo-

del with a large number of descriptors for smaller datasets, because the initial mapping of objects into a plane in fact projects the n -dimensional vectors into 2D plane. The reduction of variables is nevertheless preferable, because the robustness of the model decreases with increasing number of variables. The predicted pK_i value was 4.03, the difference between experimental and predicted value thus being 1.29.

The set of 55 compounds (see Table 1) including nucleobase, nucleosides, nucleotides and various endogenous compounds and drugs was treated separately, following the same modelling methodology as for the flavonoids. The resulting neural network model was able to predict the pK_i values of compounds from the training and testing set with a RMSEP of 0.51 and 0.26, respectively. Only three compounds of the available external validation set had been already tested experimentally, and the RMSEP of these compounds was 0.47. In Figure 2 the predictions of the training, test and validation set are shown in comparison with the experimental data. Molecular des-

criptors selected in the optimized model are given in Table 4.

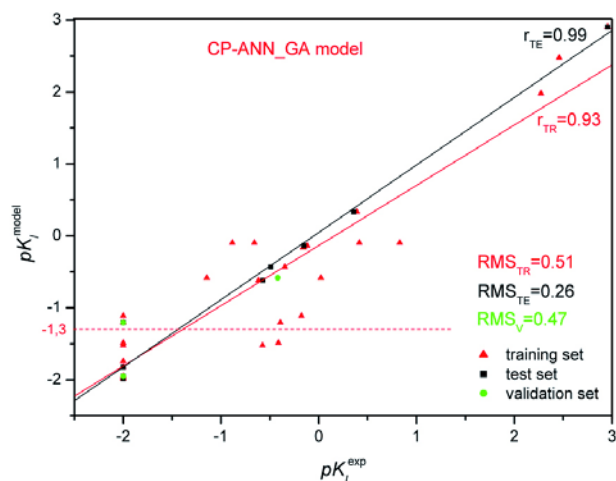


Figure 2. Regression lines of the experimental versus predicted inhibition constants pK_i , obtained by CP-ANN model coupled with GA for training (red triangles) and test (black squares) compounds, and the predictions of three compounds (green diamonds) from the external validation set.

Table 4. Selected descriptors with genetic algorithm (GA) for the counter-propagation artificial neural network model of purine derivatives.

Selected descriptors	
<i>Constitutional:</i>	
1	Number of C atoms
2	Number of N atoms
3	Number of double bonds
4	Relative number of double bonds
5	Relative number of aromatic bonds
<i>Topological:</i>	
6	Information content (order 1)
7	Complementary Information content (order 0)
8	Average Complementary Information content (order 1)
9	Average Bonding Information content (order 0)
<i>Electrostatic:</i>	
10	FPSA-2 Fractional PPSA (PPSA-2/TMSA) [Zefirov's PC]
11	RPCG Relative positive charge (QMPOS/QTPLUS) [Zefirov's PC] ³
12	RNCG Relative negative charge (QMNEG/QTMINUS) [Zefirov's PC]
<i>Quantum-chemical:</i>	
13	HASA-2 [Semi-MO PC]
14	HA dependent HDSA-2/SQRT(TMSA) [Semi-MO PC]

Among selected influential descriptors are those describing the ability of molecules to form hydrogen bonds, such as »Area-weighted surface charge of hydrogen bonding acceptor atom« (HASA-2), and »HA dependent area-weighted surface charge of hydrogen bonding donor atom (HDSA-2)«. Other significant molecular des-

criptors selected describe the shape and compactness of molecules (Information content), size and distribution of positive and negative atomic charge, and also a simple constitutional descriptor, *i.e.* the number of double bonds. The former group of descriptors associated with the ability of hydrogen bond formation support the hypothesis that the molecules, which are active as inhibitors of the bilitranslocase, pass the cell membrane by forming reversible H-bonds to the amino acid sequences of bilitranslocase positioned within the cell membrane. The same observation was explained on the set of flavonoids [2]. The weak H-bonds might enable passing the molecule by attaching and detaching dynamically to the membrane alpha helices of the studied transmembrane protein. The latter group of descriptors reflects the non-even charge distribution, or in other words, ionic properties of the molecules. A negative correlation can be found between the relative negative charge (of the molecule) and inhibition activity of tested molecules, Figure 3, which indicates that the capacity of bilitranslocase to transport ionic species and is in agreement with previous experiments.^{11,12}

4. Conclusions

From the model developed in *Part I* we can conclude that a reliable prediction (estimated error is below 10% based on external validation) of transmembrane alpha helices can be obtained for structurally diverse transmembrane proteins. The prediction of four transmembrane segments in BTL is not yet confirmed experimentally, however, it is not in conflict with the available experimental data.

In *Part II* it was found that interactions between bilitranslocase and small molecules rely on the ability to establish hydrogen bonds, diminishing the involvement of charge interactions. The results of this work show that, contrary to dietary anthocyanins, most of dietary flavonols do not interact with bilitranslocase, whereas, some flavonol aglycones act as poor ligands of that carrier. In case of nucleobases and their derivatives (nucleotides and nucleosides) the phosphate group in principle improved the transport ability by bilitranslocase. The analysis of resulting models revealed that the hydrogen bonding ability was the main information content in the selected chemical descriptors. The quantitative analysis of the structure–activity relationship led to the identification of parts of ligands potentially involved in the binding to bilitranslocase, along with a reliable hypothesis on the kind(s) of interaction between the ligand and the target. The details about specific amino acid residues of bilitranslocase involved in the interactions of passing small molecules through the membrane will be further investigated. For the time being we anticipate that the N-terminal flanking part (residue 65–75) of the second TM segment of BTL should be considered as initiator of the active transport by bili-

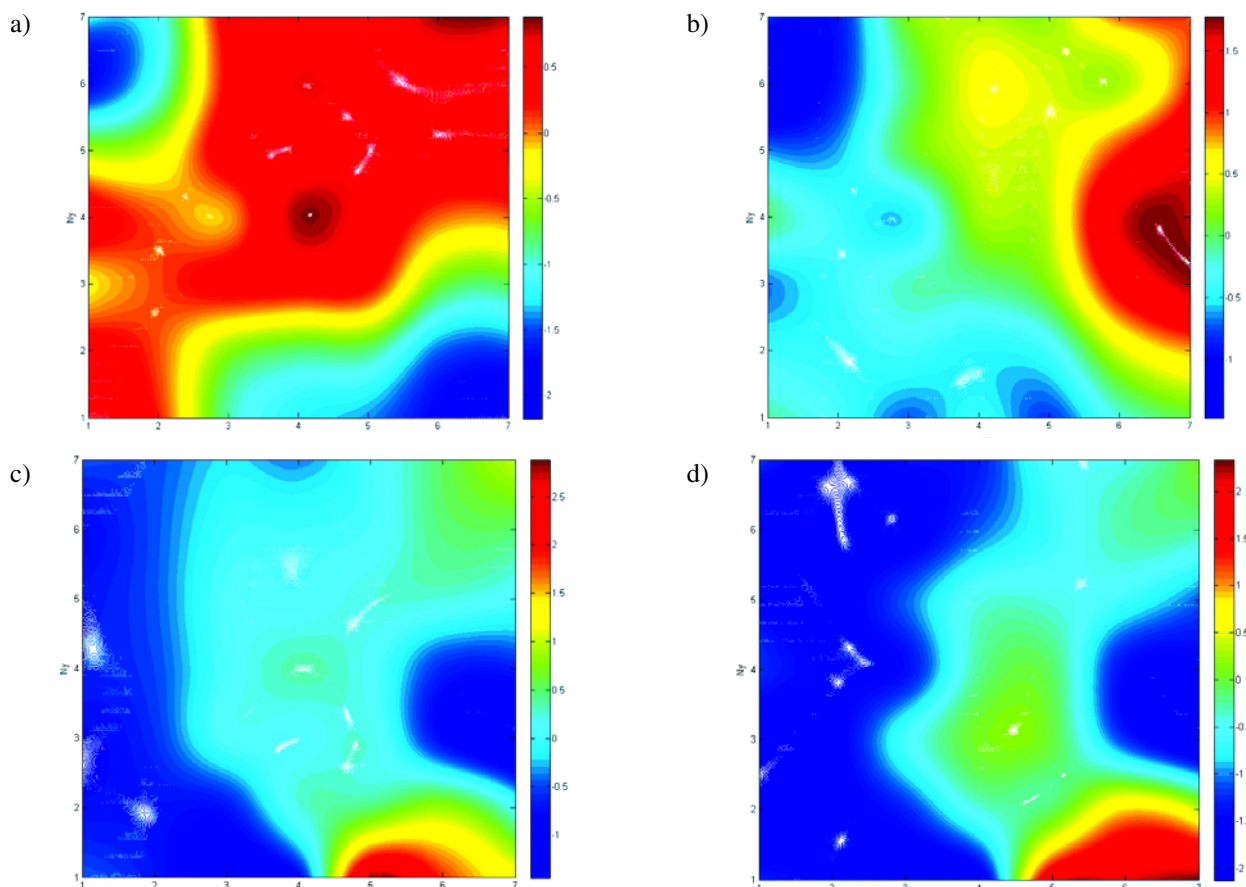


Figure 3. Weight maps of CP-ANN model related to three variables and the output layer: HA dependent area-weighted surface of hydrogen bonding donor atom (a), Area-weighted surface of hydrogen bonding acceptor atom (b), number of double bonds (c), and the output layer with response surface of inhibition constants (pK_i) (d).

translocase, and the second TM segment itself is participating in forming the transport channel wall.

5. References

1. L. Battiston, S. Passamonti, A. Macagno, G.L. Sottocasa, *Biochem. Biophys. Res. Commun.* **1998**, *247*, 687–692.
2. A. Karawajczyk, V. Drgan, N. Medic, G. Oboh, S. Passamonti, M. Novič, *Biochem. Pharm.* **2007**, *73*, 308–320.
3. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucl. Acids Res.* **2000**, *28*, 235–242, Database available at <http://www.pdb.org/pdb/home/home.do>.
4. G. E. Tusnady, Z. Dosztanyi, I. Simon, *Bioinformatics* **2004**, *20*(17), 2964–2972, Database available at <http://pdbtm.enzim.hu/>.
5. J. Zupan, M. Novič, I. Ruisanchez, *Chemom. Intell. Lab. Syst.* **1997**, *38*, 1–23.
6. J. Zupan, J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, Wiley-VCH: Weinheim, Germany, **1999**.
7. Perl 5.10.1 The Perl Foundation, **2007**; Software available at <http://www.perl.org/>.
8. Š. Župerl, S. Fornasaro, M. Novič, S. Passamonti, *Anal. Chim. Acta* **2011**, Accepted manuscript. Doi:10.1016/j.aca.2011.07.004.
9. A. R. Choudhury, M. Novič, *SAR QSAR Environ. Res.* **2009**, *20*, 741–754.
10. A. R. Choudhury, M. Novič, *International Journal of Chemical Modeling*, sent for publication [Presented at the workshop Visualization and Modeling in Chemistry, October 29–31, 2010, Split, Croatia]
11. S. Passamonti, M. Terdoslavich, R. Franca, A. Vanzo, F. Tramer, E. Braidot, E. Petrusa, A. Vianello, *Curr. Drug Metab.* **2009**, *10*, 369–394.
12. S. Passamonti, M. Terdoslavich, A. Margon, A. Cocolo, N. Medic, F. Macri, G. Decorti, M. Franko, *FEBS J.* **2005**, *272*, 5522–5535.
13. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
14. A. R. Katritzky, V. S. Lobanov, M. Karelson, *Codessa 2.0, Comprehensive Descriptors for Structural and Statistical Analysis*, University of Florida, U.S.A., **1996**.
15. R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*. Wiley-VCH, Weinheim **2009**.
16. T. Kohonen, *Overture, Self-Organizing Neural Networks*:

- Recent Advances and Applications, Springer-Verlag, Inc. New York, 2001.
17. P. Gramatica, P. Pilutti, E. Papa, *J. Chem. Inf. Comp. Sci.* 2004, 44, 1794–1802.
18. L. Davis, Handbook of Genetic Algorithms, Van Nostrand Reinhold, New York, 1991.

Povzetek

Predstavljen je pristop k analizi kemijske strukture bilitranslokaze, membranskega proteina, ki prenaša bilirubin iz krvi v jetrne celice. Zaporedje aminokislin in informacije o sekundarni strukturi transmembranskih segmentov proteinov s poznano 3D strukturo uporabimo za napovedovanje transmembranskih domen strukturno nerešenih proteinov. S pomočjo poznanih struktur kodiramo transmembranske in ostale domene proteinov na tak način, da jih lahko s pomočjo računalniških programov grupiramo glede na njihove podstrukturne karakteristike in da lahko zgradimo model za napovedovanje transmembranskih segmentov. Prikazan model za napovedovanje transmembranski segmentov identificira štiri transmembranske alfa vijačnice, od katerih vsaka vsebuje okoli 20 aminokislin. Ta rezultat je delno potrjen z eksperimentalnimi študijami uporabe določenih protiteles, ki ustrezajo delom aminokislinskega zaporedja bilitranslokaze. Da bi razjasnili transportni mehanizem bilitranslokaze, smo testirali tudi niz strukturno raznolikih spojin za določitev njihove kompetitivne konstante inhibicije v preiskovanem sistemu protein-substrat. Informacija o kemijski strukturi majhnih molekul, ki ali prehajajo ali blokirajo transmembransko pot preko bilitranslokaze, nam pomaga graditi hipotezo o transportnem mehanizmu študirane biološkega sistema.