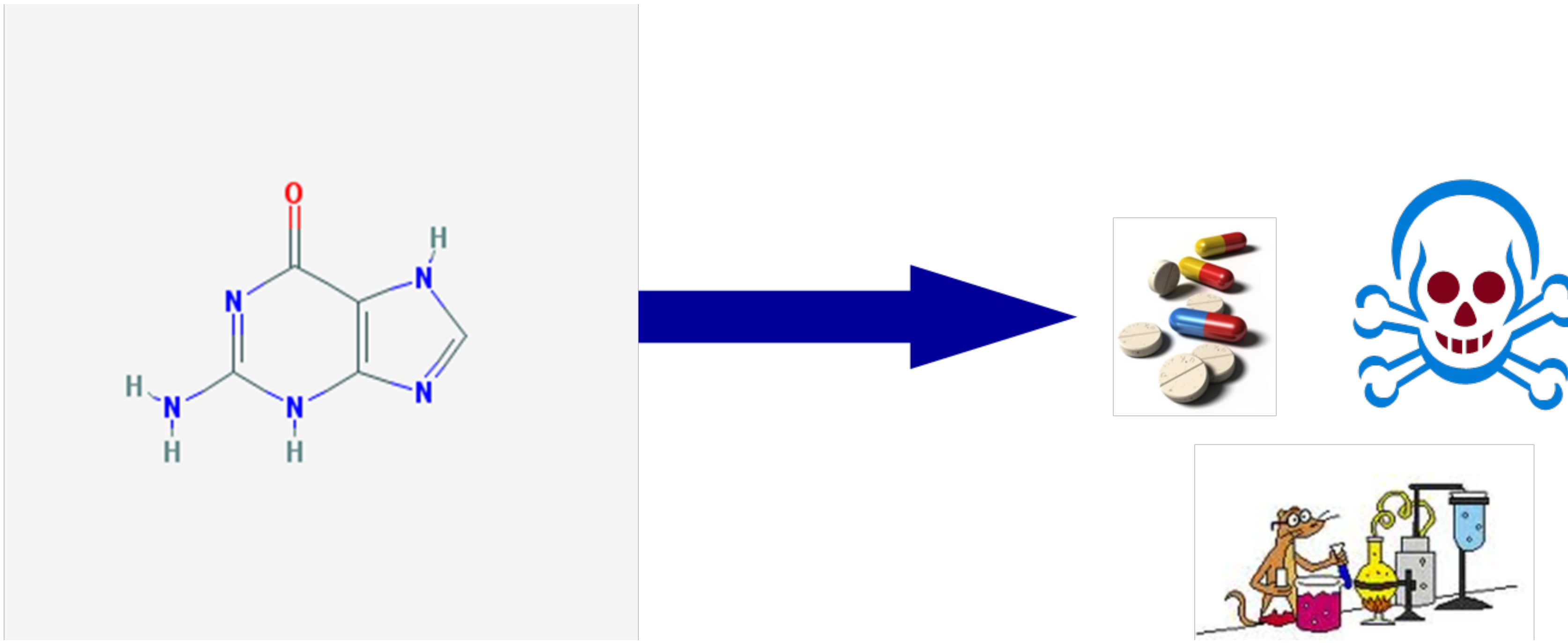
Towards Artificial Intelligence by Artificial Neural Networks

Dr. Marjana Novič
Laboratory for Cheminformatics

National Institute of Chemistry

QSAR (Quantitative Structure Activity Relationship)



QSAR (Quantitative Structure Activity Relationship)

DIFFERENT
TOOLS

Computational methods
Mathematics
Statistics

Exploitation of different tools in the research and application of uni- and multi-variate problems in chemistry

DIFFERENT TYPES

of uni-variate and multi-variate DATA

Origin: problems in chemistry

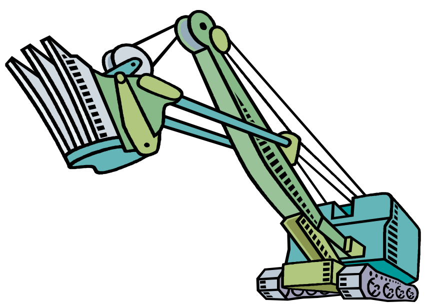
QSAR

DIFFERENT TOOLS

A tool has to be adapted to DATA

Shoveling is a difficult task for which a shovel is needed

However, for shoveling different materials, different shovels are required.

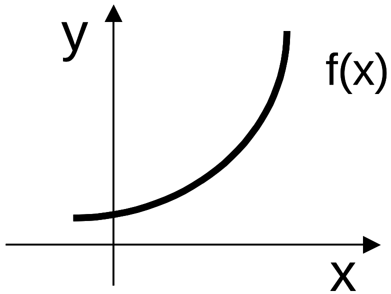


QSAR- Tools

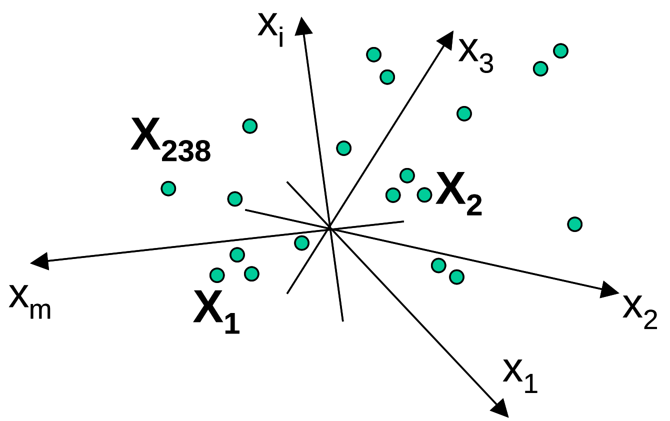
The same task (shoveling or data handling) must be carried out by different tools if different material (data) has to be handled most efficiently.

QSAR- Data

DIFFERENT TOOLS / DATA



$y = f\ (x, \text{parameters})$



$Y(y_1, y_2, \ldots y_n) = \text{MODEL}\ [X(x_1, x_2, \ldots x_m), \text{parameters}]$

QSAR- Data

DIFFERENT TOOLS

DATA

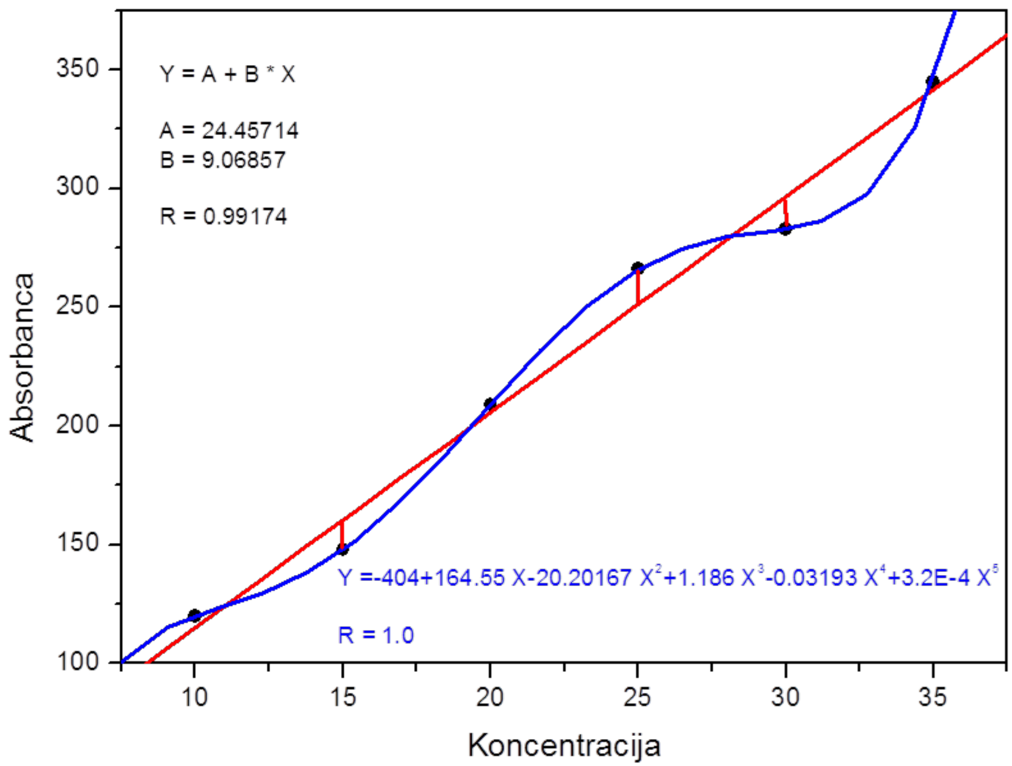
Pattern or"chemical image" of the analysed sample denotes $X=(x_1, x_2, \ldots x_n)$.

Sample's components are features obtained by measurements and specified by the specialist.

Chemometrician's role is to choose particular variables obtained from measurements to enable correct description, classification...

Each sample is characterised by a unique typical set of data, "fingerprint" in m-dimensional pattern space.

QSAR- Tools



X	Y
10	120
15	148
20	209
25	266
30	283
35	345

QSAR- Data

DIFFERENT TOOLS

DATA

Genomics and proteomics are two directions that are aplicable also in food authentication.

Large amount of data demand specific treatment and tools.

Graph-theoretical approach has shown pro-missing results in genomics and proteomics.

Genomics

RANDIĆ, M. Graphical representations of DNA as 2-D map. Chem. Phys. Lett., 2004, 386, 4/6, 468-471.
RANDIĆ, M, VRAČKO, M, ZUPAN, J, NOVIČ, M. Compact 2-D graphical representation of DNA. Chem. Phys. Lett., 2003, 373, 5/6, 558-562.

Proteomics

RANDIĆ M, NOVIČ M, VRAČKO M. Novel characterization of proteomics maps by sequential neighborhoods of protein spots, *J. chem. inf. model.*, 2005, 45, 1205-1213.
RANDIĆ, M, ZUPAN, J, NOVIČ, Ma. On 3-D graphical representation on proteomics maps and their numerical characterization. *J. chem. inf. comput. sci.*, 2001, 41, 1339-1344.
RANDIĆ, M. On graphical and numerical characterization of proteomics maps. *J. chem. inf. comput. sci.*, 2001, 41, 1330-1338.

QSAR- Tools

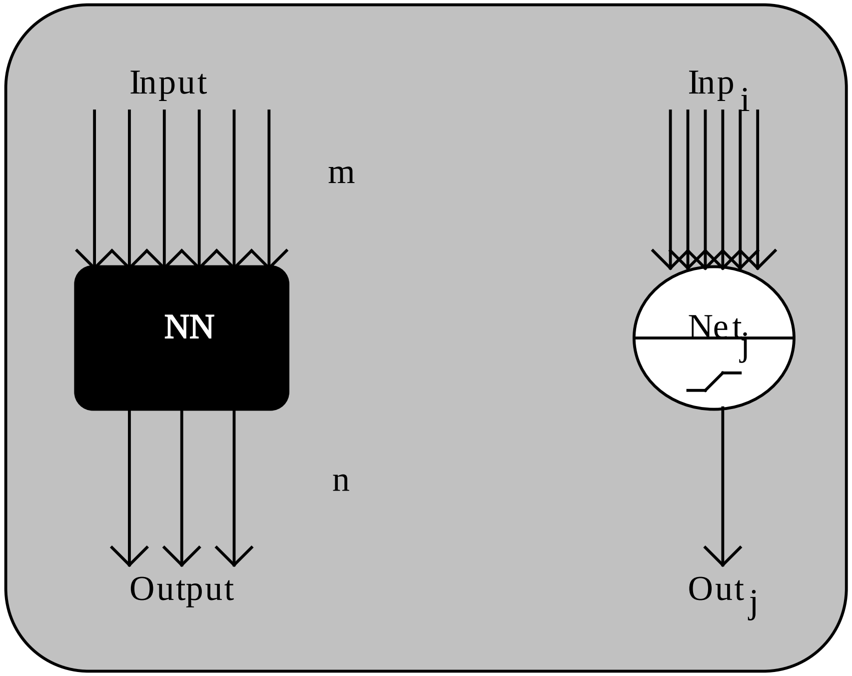
DIFFERENT TOOLS

DATA

Descriptive statistics
Predictive statistics
Expert systems
Pattern recognition
Artificial intelligence
Calibration
Signal processing
Regression methods
Neural networks
Experimental designs
Optimisations, etc...

Machine learning – Artificial neural networks (ANN)

ANN as a black box for making decisions (pattern recognition, determination of various features, process control...)



Input: X=(x_1,x_2,...x_i,...x_m)

ANN becomes useful when it is properly **trained** for a desired purpose.

Output: Y=(y_1,y_2,...y_j,...y_n)

Machine learning – Artificial neural networks (ANN)

ANN as a black box for making decisions
Applications: tasks that are hard to solve using ordinary rule-based programming, including Computer Vision and Speech Recognition, Character Recognition, Image Compression, Stock Market Prediction, Traveling Saleman's Problem, Problems in Medicine, Life Sciences, Electronic Nose, Security, and Loan Applications

- **Architecture** (The interconnection pattern between the different layers of neurons)
- The **learning process** for updating the weights of the interconnections
- The **activation function** that converts a neuron's weighted input to its output activation.

Types of ANNs

- Division with respect to the way of training
  - Supervised (1)
  - Unsupervised (2)

- (1) Error back propagation ANN – based on perceptron
- (1) Radial basis function RBF ANN
- (2) Kohonen ANN
- (2..1) Counterpropagation ANN

ANNs

Supervised learning:

The relationship between objects and targets is known in advance for a set of objects (training set).

$$Y=f(X)$$

Object

$X=(x_1,x_2,\ldots x_i,\ldots x_m)$

Target

Y=(known property)

Supervised ANNs

1943: McCulloch in Pitts, "A Logical Calculus of Ideas Immanent in Nervous Activity"

1949: Donald Hebb, "The Organization of Behavior", Hebb rule (strengthning connections between neighbouring fired neurons – associated with memorizing)

1962: Frank Rosenblatt, "Principles of Neurodynamics" McCulloch-Pitts- nervus activity and Hebb rule: **perceptron**.

1969: Minsky & Papert, "Perceptrons" Theoretical limitations in nonlinear problems

1982: John Hopfield, nonlinearity introduced in neurons.

1986: Rumelhart, Hinton, Williams: Learning representations by back-propagating errors

letters to nature

Nature **323**, 533 - 536 (09 October 1986); doi:10.1038/323533a0

Learning representations by back-propagating errors

DAVID E. RUMELHART[*], GEOFFREY E. HINTON[†] & RONALD J. WILLIAMS[*]

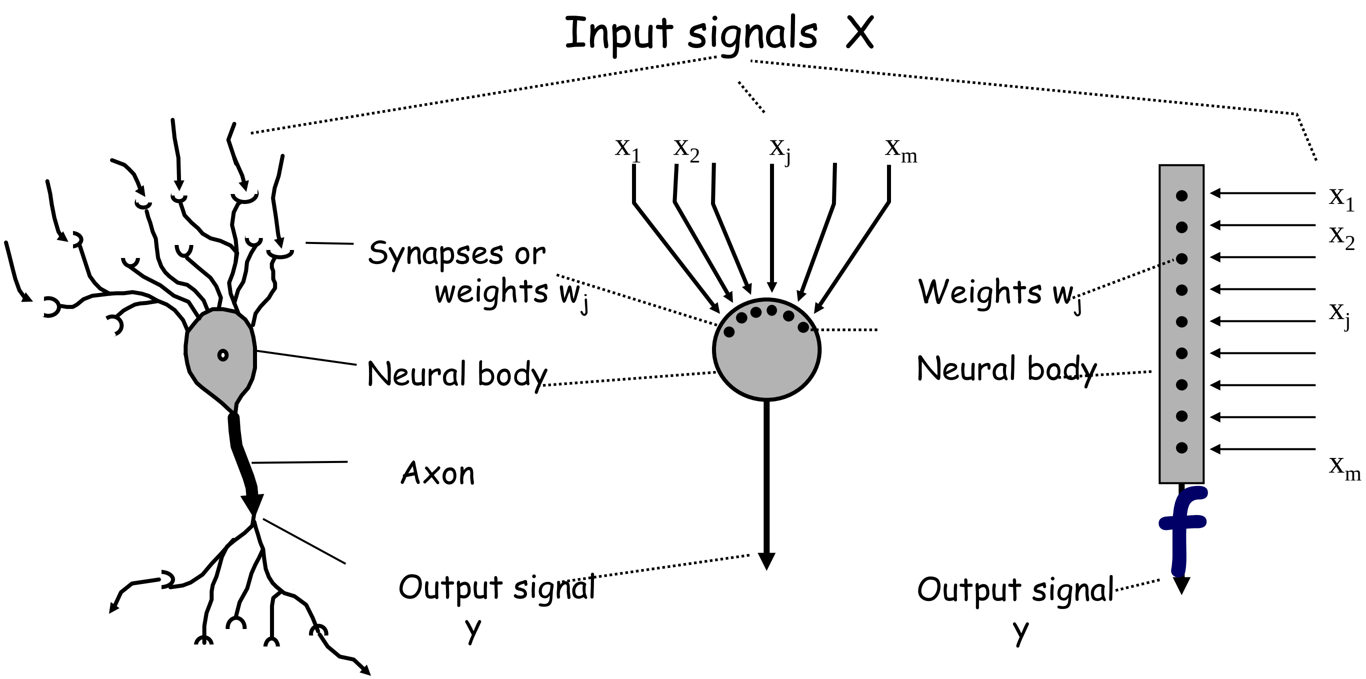
[*]Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA
[†]Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA
[†]To whom correspondence should be addressed.

We describe a new learning procedure, back-propagation, for networks of neurone-like units. The procedure repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector. As a result of the weight adjustments, internal 'hidden' units which are not part of the input or output come to represent important features of the task domain, and the regularities in the task are captured by the interactions of these units. The ability to create useful new features distinguishes back-propagation from earlier, simpler methods such as the perceptron-convergence procedure[1].

References

1. Rosenblatt, F. *Principles of Neurodynamics* (Spartan, Washington, DC, 1961).
2. Minsky, M. L. & Papert, S. *Perceptrons* (MIT, Cambridge, 1969).
3. Le Cun, Y. *Proc. Cognitiva* **85**, 599−604 (1985).
4. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Vol. **1**: *Foundations* (eds Rumelhart, D. E. & McClelland, J. L.) 318−362 (MIT, Cambridge, 1986).

ANNs : Neurons



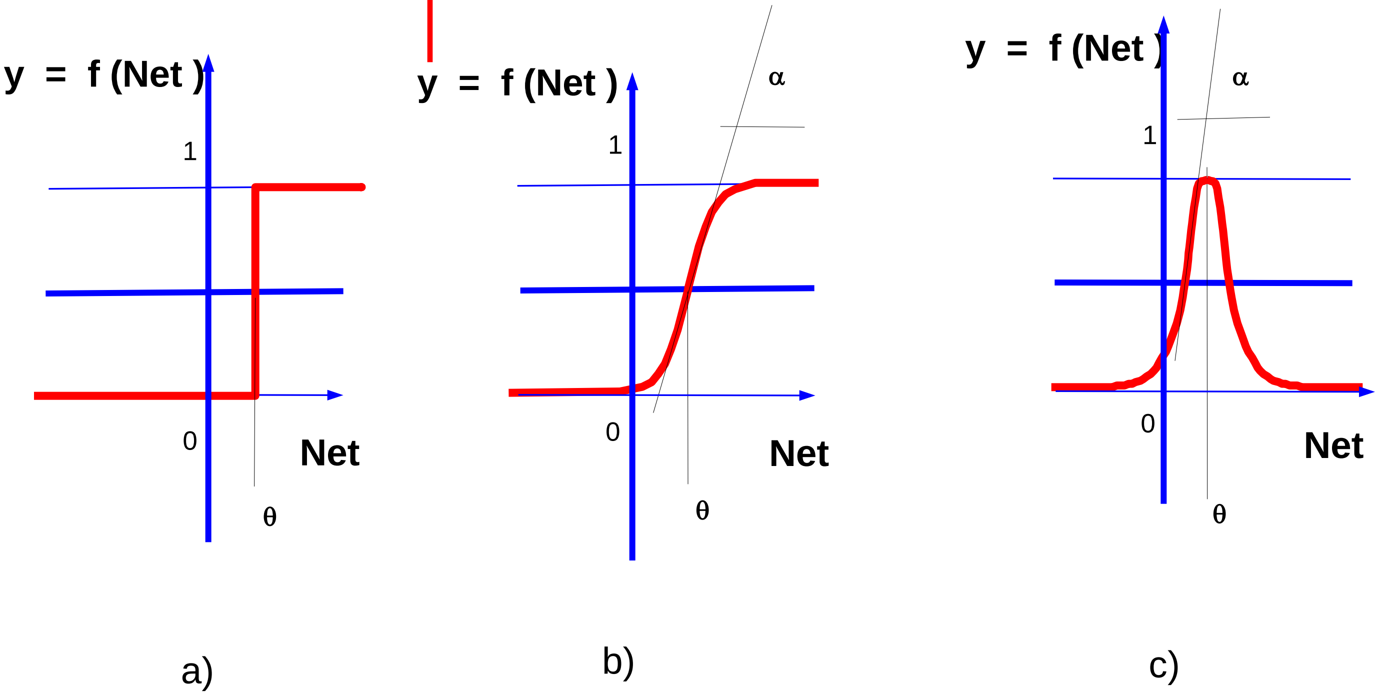
Output signals of neurons

"Error Back Propagation" ANNs

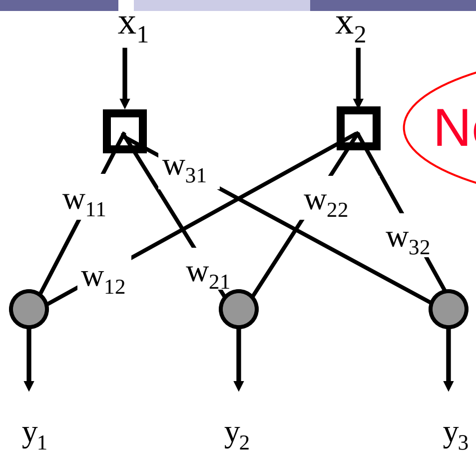
$Net_j = \Sigma_{i=1,k}\, w_{ji}\, x_i$

$$y = \frac{1}{1+e^{-\alpha(Net-\theta)}}$$

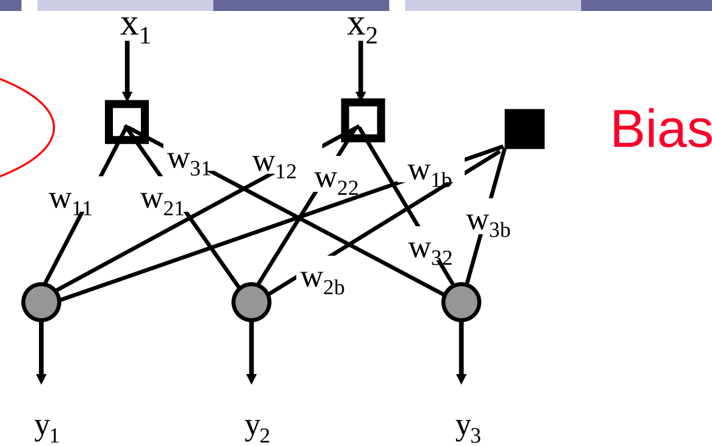
$$Y = signal = out$$

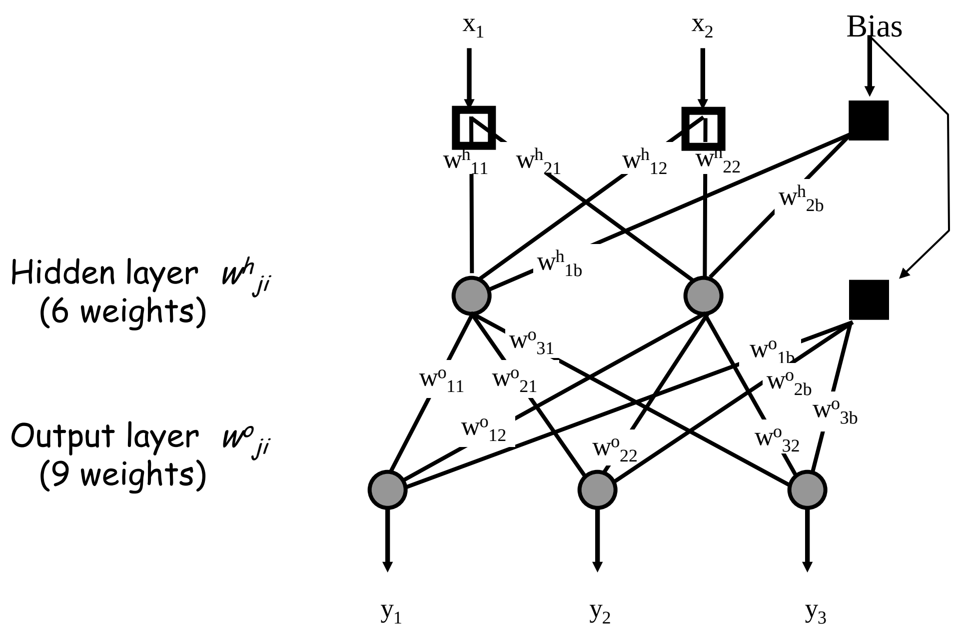


Various architecture of ANNs Error-back-propagation



$$Net_j = \Sigma_{i=1,k}\, w_{ji}\, x_i$$





Learning: Correction of weights

$$\Delta w_{ji}^{l} = \eta \delta_j^{l} out_j^{l-1} + \mu \Delta w_{ji}^{l,prejšnji}$$

$$signal = out$$

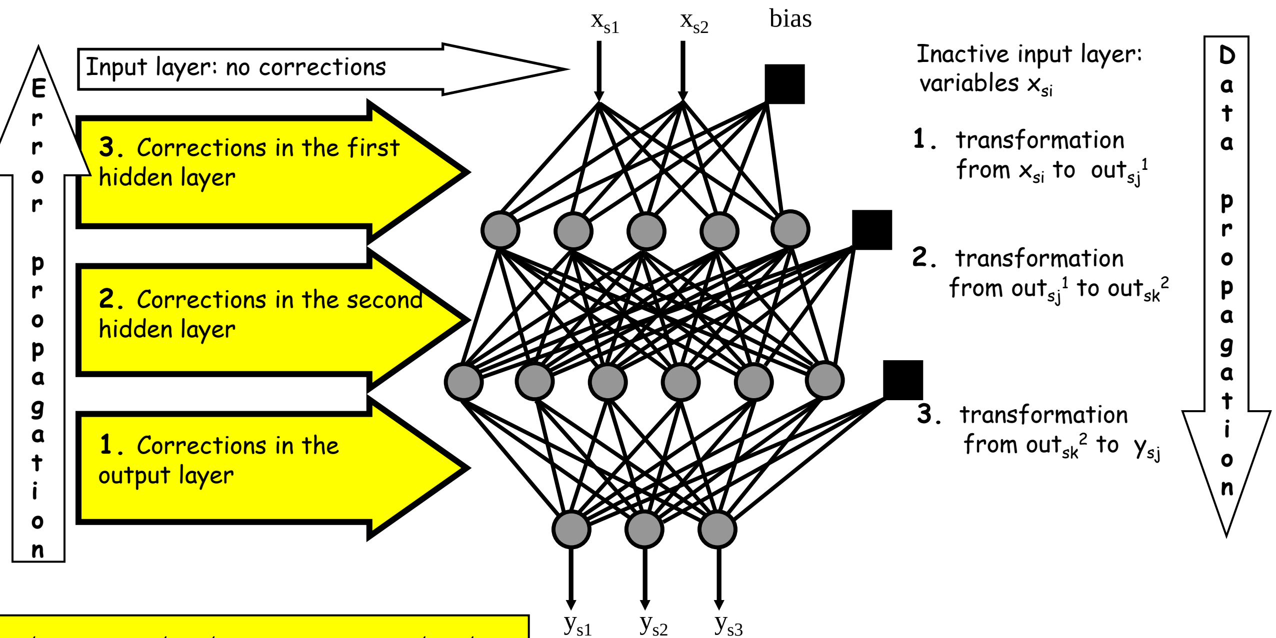
$$signal = \frac{1}{1+e^{-\alpha(Net-\theta)}}$$

$$Net_j = \Sigma_{i=1,k} w_{ji} x_i$$

$$\delta_j^{hidden} = (\sum_{k=1}^{n_r} \delta_k^{output} w_{kj}^{output}) y_j^{hidden} (1 - y_j^{hidden})$$

Learning rate [0,0..1,0]

momentum



$$\delta_j^{output} = (t_j - y_j^{output}) y_j^{output} (1 - y_j^{output})$$

ANNs

ANN Optimization

Minimization of the objective function
(mean square error function: loss/cost function)

Gradient based search techniques (problem: local opt.)
Simulated Annealing (Global opt.)
Genetic Algorithm (Global opt.)

ANNs

Unsupervised learning:

The relationship between objects and targets is not defined in advance.

Object $X=(x_1,x_2,\ldots x_i,\ldots x_m)$

Target Y=(not known)

Unsupervised ANNs

Tuevo Kohonen (born 1934), Finnish academician. “Self-organizing maps”

1960: T. Kohonen : “neural computing” associative memory, optimal associative mapping, self-organizing feature maps (SOMs).

2007: T. Kohonen, and T. Honkela: "Kohonen network“

http://www.scholarpedia.org/article/Kohonen_network

Kohonen - ANN = Self Organizing Maps (SOM)

Basic principles of Kohonen learning

- organization of neurons
    - 1-dimensional (in a line)
    - 2-dimensional (in a map)
- analogy of Kohonen NN with brains
- learning strategy - "winner takes all"
- learning rate parameter
- amount and area of weight correction
    - shrinking of the neighbourhood
- recognition of objects from training-set
- visualization of clusters formed in the top-map

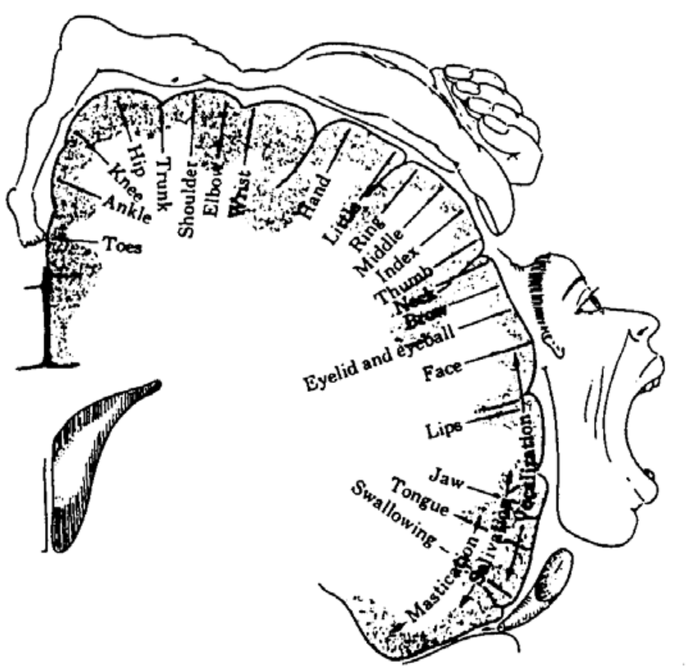
Kohonen - ANN (SOM)

Basic principles of Kohonen learning

- prediction about "unknown" objects
- top-layer(TL) of labels
  - empty spaces in the TL
  - clusters in the TL
  - conflicts in the TL
- weights levels

Kohonen - ANN (SOM)

Kohonen learning can be used for a projection of **multi-dimensional** objects into a **two**-dimensional plane (map)

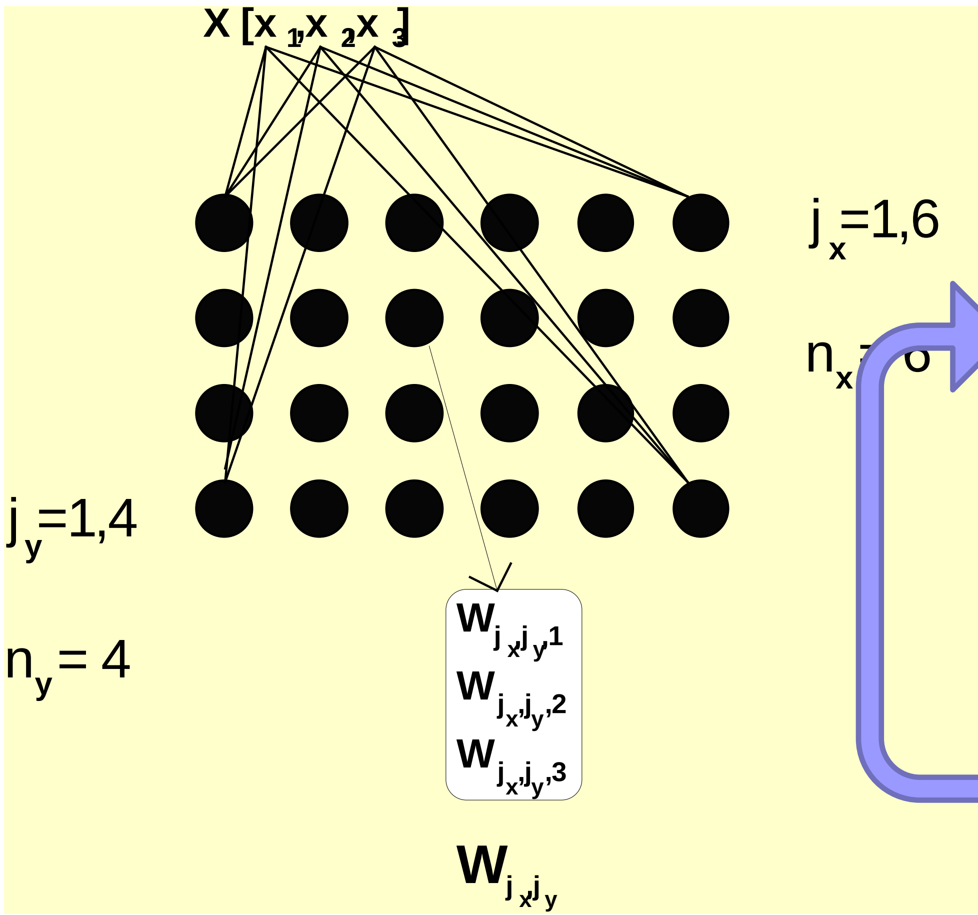


"**homunculus**"

The term appears to have been first used by the alchemist Paracelsus

Kohonen - ANN (SOM)

2-dimensional organization of neurons (in a map)



Number of neurons

$n = n_x n_y = 24\ (6 \times 4)$

X $[x_1, x_2, x_3]$

W $[w_1, w_2, w_3]$

Euclidean distance at one neuron W

$$E_D = \sqrt[2]{\sum_{i=1}^{m} (X_i - W_i)^2}$$

Kohonen - ANN (SOM)

Learning strategy - "winner takes all"

Winner = W_{win} <---min (ε_j)

$$\varepsilon_j = \sum_{i=1}^{m} (x_i - w_{ji})^2$$

x_i <-- object-components
w_{ji} <-- j^{th} neuron-components (weights)
ε_j <-- difference between the object and the j^{th} neuron

Kohonen - ANN (SOM)

CORRECTION:

"winner" neuron + surrounding neurons

learning rate parameter **η(t,r(t)))**

t: time of learning
r(t): neighbourhood

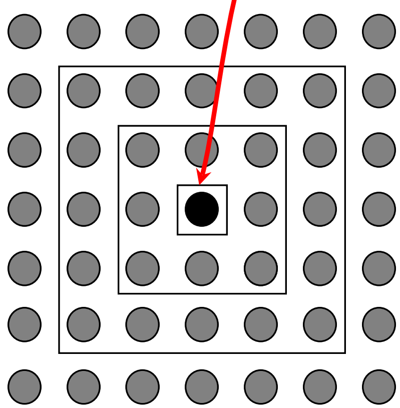
$$w_{ji}^{new} = w_{ji}^{old} + \eta\left(x_{si} - w_{ji}^{old}\right)$$

$$i = 1, m$$

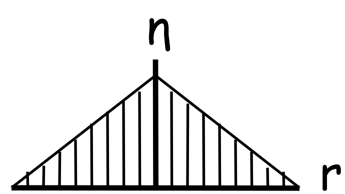
$$j = 1, n$$

$$s = 1, p$$

Kohonen - ANN (SOM)



The correction of the "winner" and the surrounding neurons for one (sth) object is regarded as **one iteration step**



Triangular function η=f(t, r(t))

t: training time
r(t): neighbouring neurons

When all objects went through the network, the whole procedure, called one epoch, is repeated

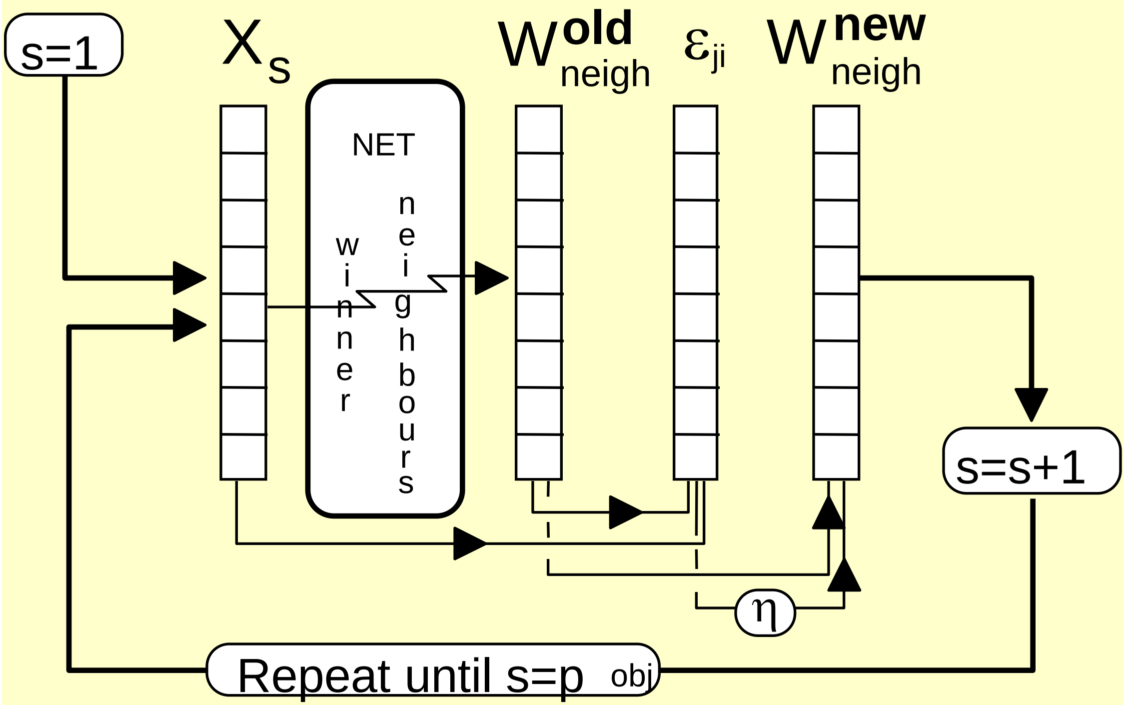
p iteration steps = **1 epoch**

Kohonen - ANN (SOM)

One epoch of Kohonen learning

$$\varepsilon = X_s - W^{\mathbf{old}}_{neigh}$$

$$w_{ji}^{new} = w_{ji}^{old} + \eta(x_{si} - w_{ji}^{old})$$



p : number of all objects in the training set

Kohonen - ANN (SOM)

Recognition of objects from the training set

When the training is completed, all the objects from the training set must be recognizable

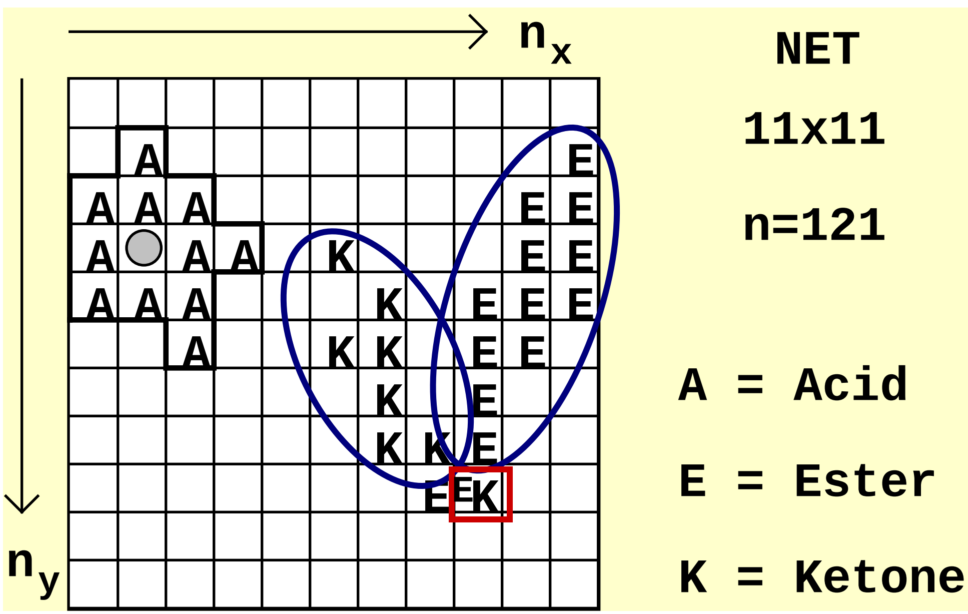
Usually the above requirement is achieved when the total error of one epoch: ε_{epoch} is below a specified limit

$$\varepsilon_{epoch} = \sum_{s=1}^{p} \sum_{j=1}^{n} \sum_{i=1}^{m} (x_{si} - w_{ji})^2$$

Kohonen - ANN (SOM)

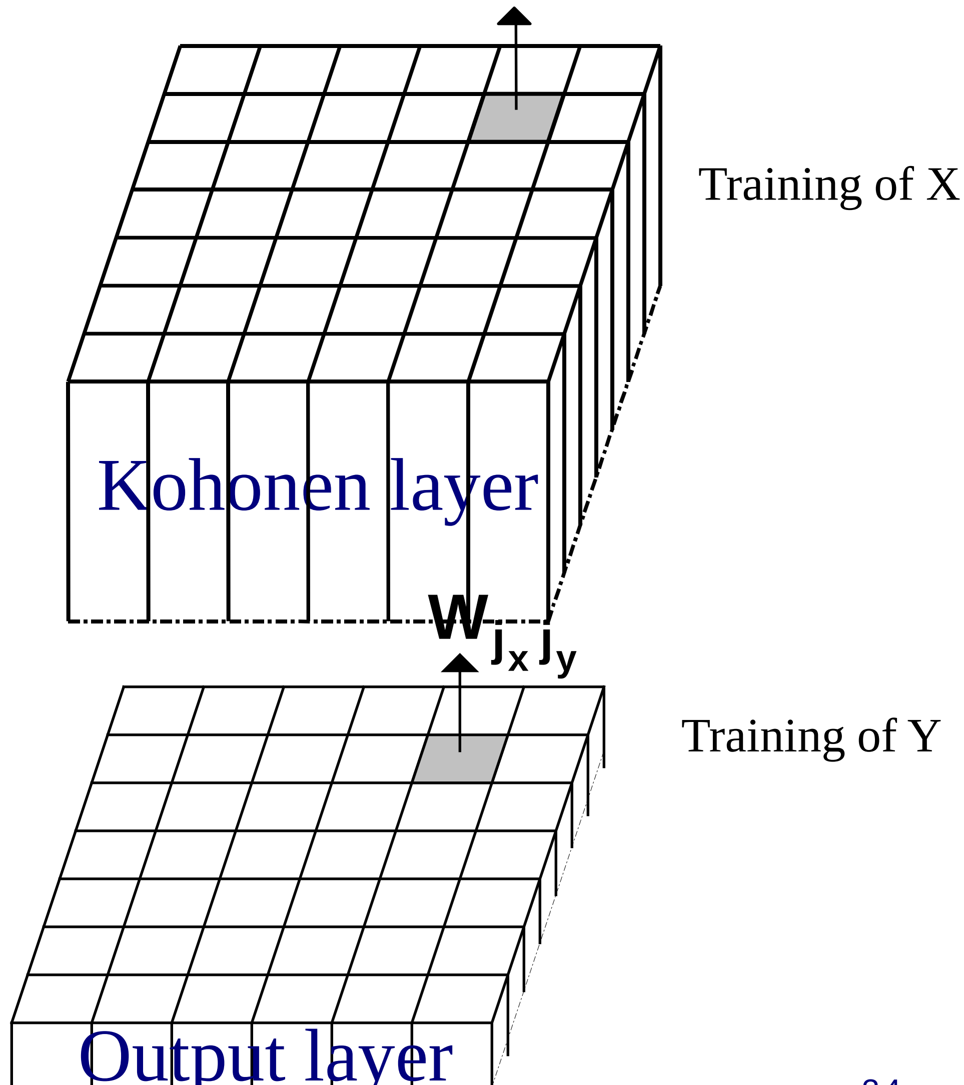
Top-layer (TL) of labels:

-empty spaces in the TL
-clusters in the TL
-conflicts in the TL



Counter-propagation - ANN

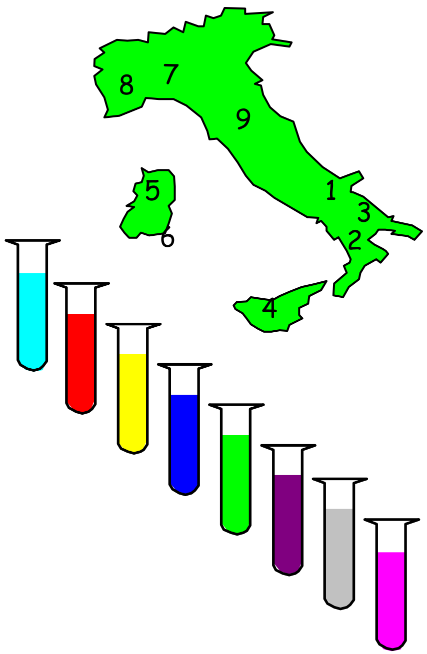
$$w_{ji}^{new} = w_{ji}^{old} + \eta(x_{si} - w_{ji}^{old})$$
$$i = 1, m$$
$$j = 1, n$$
$$s = 1, p$$



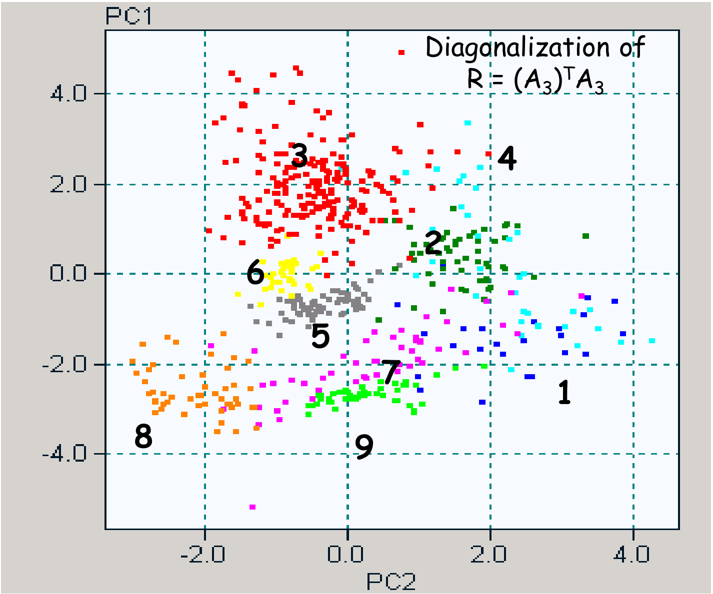
$$w_{ji}^{new} = w_{ji}^{old} + \eta(y_{si} - w_{ji}^{old})$$
$$i = 1, t$$
$$j = 1, n$$
$$s = 1, p$$

(Q)SAR CASE 1

Analysis of 572 olive oils from 9 Italian regions



PCA score plot

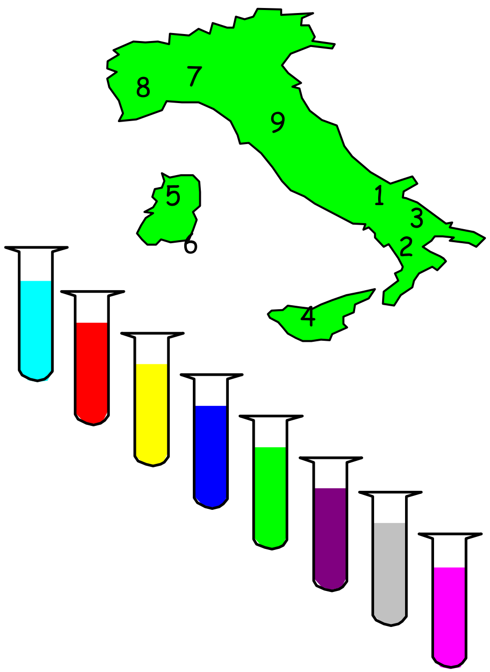


M. Forina and C. Armanino, Ann. Chim. (Rome), 72 (1982) 127.
M. Forina and E. Tiscornia, Ann. Chim. (Rome), 72 (1982) 144.

8 fatty acids concentration

(Q)SAR CASE 1

Analysis of 572 olive oils with Kohonen ANN
Unsupervised learning



1 North Apulia 25
2 Calabria 56
3 South Apulia 206
4 Sicily 36
5 Inner Sardinia 65
6 Coastal Sardinia 33
7 East Liguria 50
8 West Liguria 50
9 Umbria 51
Σ572

6	6	6	6	6	5	5	5	5	5	5		8	8	8	8	8	8	8	
6	6	6	6			5	5		5	5	5		7		8	8	8		8
6	6	6	6		5	5	5	5	5	5			7		8	8	8	8	8
6	6	6	6	5		5	5	5	5	5		7	7	7	7	8	8	8	8
			6	6	5	5	5	5		7		7	7	7	7	7		8	8
3	3				5	5	5		7	7	7		9	9	9	9		8	
3					5	5	5		7	7	7		9	9	9	9			
3	3					5	5		7	7	7	7	9	9	9	9	9		
3	3				3			7	7	7	7		9	9	9				1
3	3	3	3	3	3	3			7	7	7	9					1	4	1
3	3	3	3	3	3	3	2	2					4		2	1	1	1	1
3	3	3	3		3	3	3	3		2	2	2	2	2	2	1		1	1
3	3	3	3	3	3	3	3	3					2	2			1	1	1
3		3			3	3	3			3	2	2	2		2		2	4	4
3	3	3	3	3	3	3	3	3	3	3		2		2		2	2		
	3		3	3	3	3	3	3	3	3	3	2		2	2	4		4	4
3	3	3		3	3	3	3	3	3	3	3	3	4	2	2	4	4	4	4
3	3	3	3		3		3	3	3	3		4		2			2	2	2
3	3		3		3	3	3	3		3	4	4	3	4	4	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3		3	4	4	4	2	2	2	4

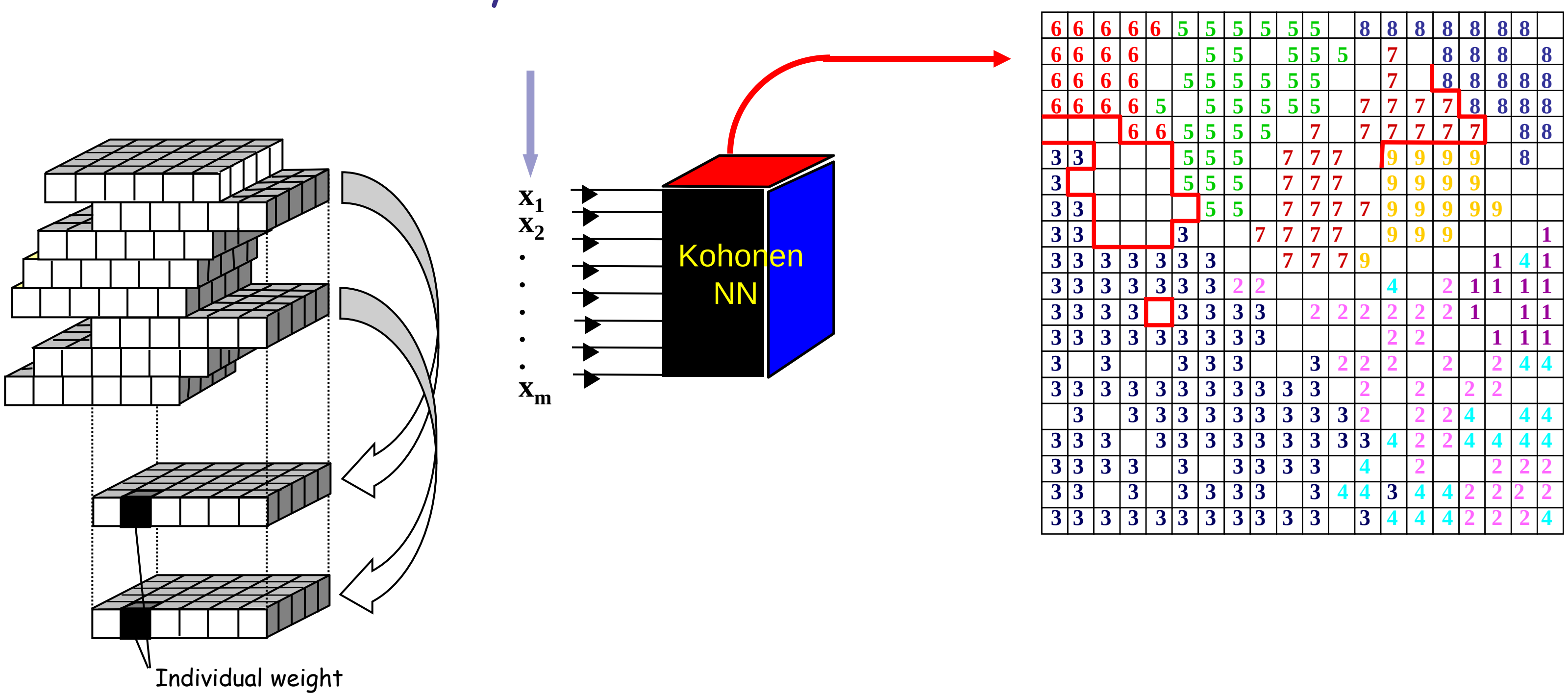
J. Zupan, M. Novic, X. Li, J. Gasteiger, Classification of Multicomponent Analytical Data of Olive Oils Using Different Neural Networks, Anal. Chim. Acta, 292, (1994), 219-234.

(Q)SAR CASE 1

Each oil is described by <u>eight fatty acid concentrations</u>

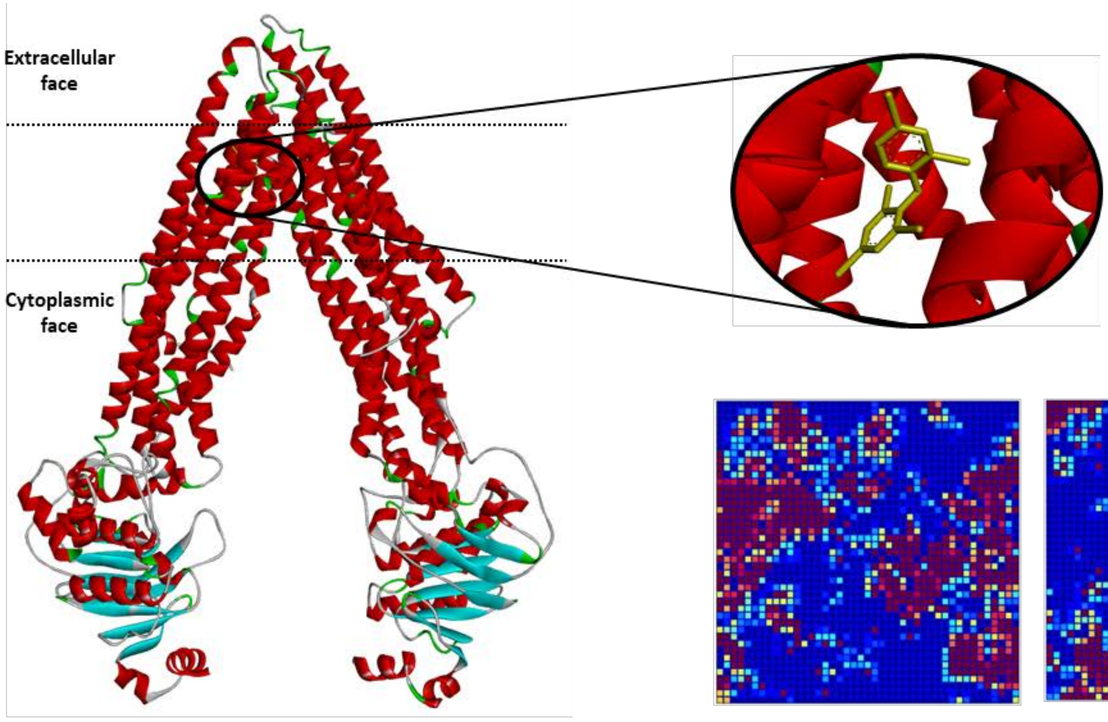
One olive oil analysis

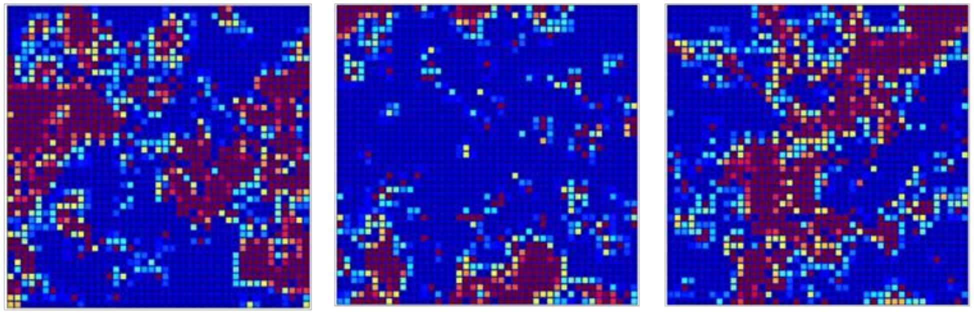
2D map of oils' origins



(Q)SAR CASE 2

P-glycoprotein (Inhibitor, Substrate, Inactive)





P-gp: Inhibitors Substrates Inactive

CONCLUSIONS – PART I

• Handling of large amounts of different data (chemical, biochemical, proteome data) require different tools

• Tools are needed for visualisation, clustering, classification, optimisation, prediction of properties...

• Missing data and values below detection limit should be properly labeled and accounted for

• Classification (of food or P-gp inhibitors) on the basis of its analysis "fingerprint" – models that learn from experience show better efficiency compared to deterministic models

Drug Design / Toxicity Assessment Based on Molecular and QSAR Modelling

Data driven models ((Q)SAR) employed as filters for molecular modelling

Marjana Novič

Laboratory for Cheminformatics
Theory Department, KI

Kemometrija je veda, ki predstavlja vez med matematiko in kemijo.

Združuje teoretične - matematične - računalniške pristope s praktičnimi aplikacijami v kemiji.

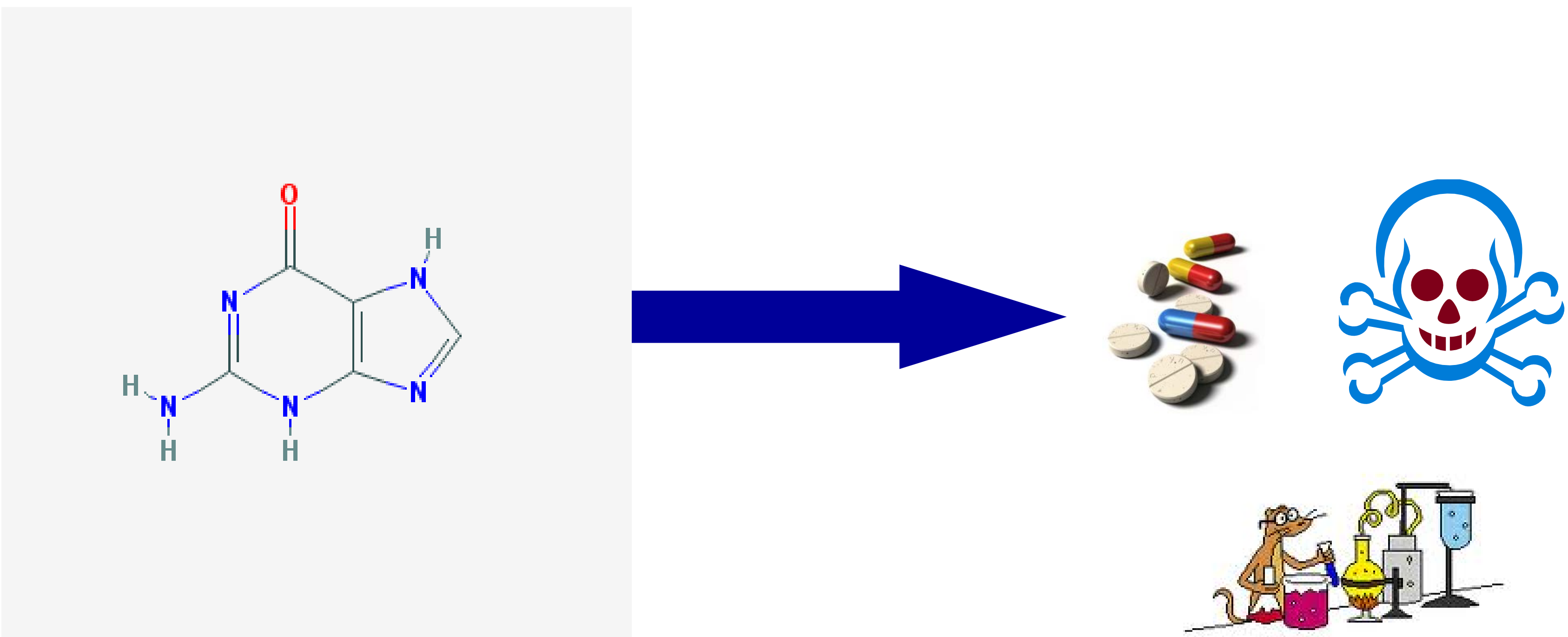
L03

Prof. dr. Marjana Novič	Znanstvena svetnica
Doc. dr. Marjan Vračko	Višji znanstveni sodelavec
Doc. dr. Marjan Tušar	Znanstveni sodelavec
Dr. Natalja Fjodorova	Znanstvena sodelavka
Dr. Katja Venko	Znanstvena sodelavka
Dr. Viktor Drgan	Znanstveni sodelavec
Dr. Nikola Minovski	Znanstveni sodelavec
Dr. Jure Borišek	Znanstveni sodelavec
Dr. Liadys Mora Lagares	Asistentka z doktoratom
Maja Kokot	MR
Eva Prašnikar	MR
Janja Sluga	MR
Benjamin Bajželj	MR
Martin Ljubič	MR

Prof. dr. Jure Zupan Zaslužni raziskovalec
Prof. dr. Milan Randić Častni član KI, 4 mesece/leto

Kemometrija – med kemoinformatiko in bioinformatiko - razvoj in uporaba metod

Raziskave temeljijo na predpostavki, da obstaja povezava med kemijsko strukturo in lastnostjo spojin



- Računalniški klaster
  - 1 rack & in-row cooling; 320 core
  - 20 računalnikov za izvedbo delavnic
- Programska oprema
  - Gaussian 09
  - Discovery Studio
  - Pipeline Pilot
  - CODESSA, DRAGON
  - QSARINS
  - CPANNatNIC (CPANN@KI)

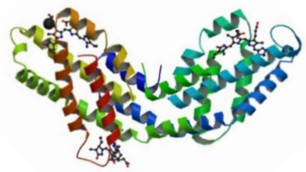
- Napovedovanje toksičnosti (preučevanje mehanizmov, ocenjevanje kemikalij) (EDs, ImageTox, Caesar, Cosmos, Prosil, In3)
- Optimizacija encimske katalize (IBAAC projekt)
- Energetika Diels-Alder reakcij (cikloadicija, ligacije BioChemLig projekt)
- Antioksidanti
- Transmembranski proteini – bilitranslokaza (T2C projekt)
- Transmembranski proteini – P-glikoprotein (In3 projekt)
- Načrtovanje inhibitorjev encimov (proteaze, hidrolaze)

Molekulsko modeliranje: Ligand (substrat)-Protein

Protein:

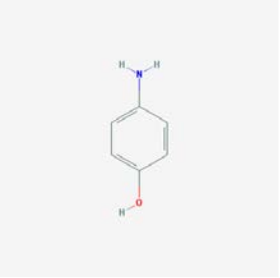
- zaporedje AA
- 3D struktura

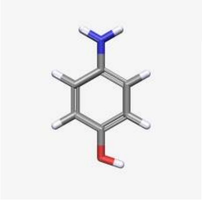
MLIHNWILTFSIFREHPSTVFQIFTKCILVSSSF........



Ligand (substrat):

- Kemijska formula $NH_2C_5H_4OH$
- 3D struktura (optimizacija)
- Izračun deskriptorjev





Podatki o strukturi in lastnostih molekul – podatkovni niz

$$Y = f(X)$$

Y

- Vezava ligand-protein
- Inhibicija proteina
- Transportna aktivnost proteina

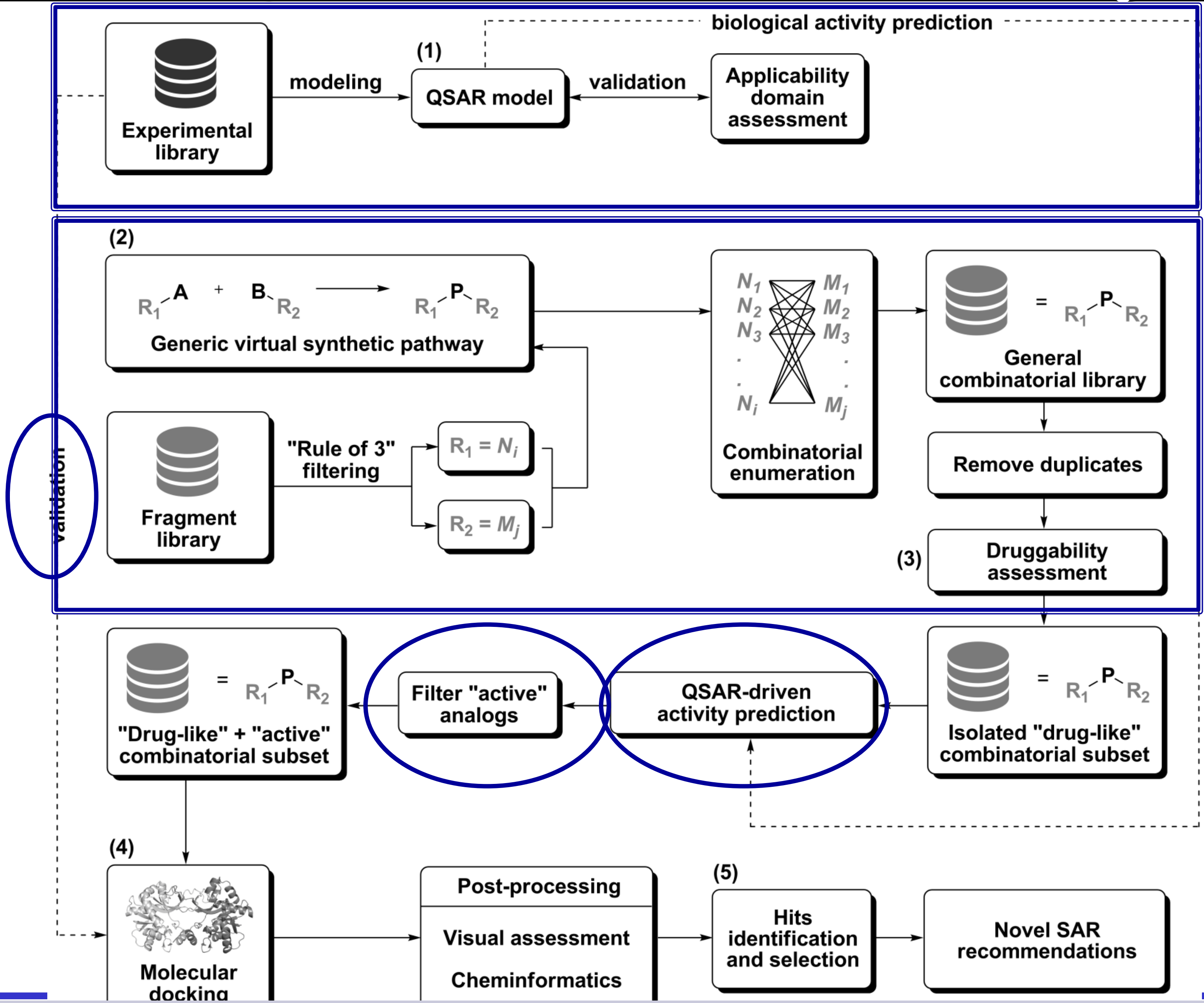
Y_i

X

Strukturni deskriptorji (m)

i-ta molekula je predstavljena kot vektor

$X_i\ (x_j,\ j=1,m)$



Minovski N, Novič M, Integrated in silico Methods for the Design and Optimization of Novel Drug Candidates: a Case Study on Fluoroquinolones – Mycobacterium tuberculosis DNA gyrase Inhibitors, Chapter in: "*Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment*", Editor: Kunal Roy, IGI Global 2014.

Bilitranslokaza

- Organski **anionski transporter** – prenašalec **Bilirubin**a iz krvi v jetrne celice
- Sekvenca poznana – **340 aminokoslin** (*Rattus norvegicus*)
- **Nima homologne struktutre v PDB**. 94% homologija nasprotne sekvence (**antisense** strand) s **ceruloplasminom**
- 3D struktura in transportni mehanizem nista poznana

MLIHNWILTFSIFREHPSTVFQIFTKCILVSSSFLLFYTLLPHGLLEDLMRRVGDSLVDLIVIC**EDSQGQHLSS**FCLFVATLQSPFSAGVSGLCKAI
LLPSKQIHVMIQSVDLSIGITNSLTNEQLCGFGFFLNVKTNLHCSRIPLITNLFLSARHMSLDLENSVGSYHPRMIWSVTWQWSNQVPAFGETS
LGFGMFQEKGQRHQNYEFPCRCIGTCGRGSVQCAGLISLPIAIEFTY**QLTSSPTC**IVRPWRFPNIFPLIACILLLSMNSTLSLFSFSGGRSGYVL
MLSSKYQDSFTSKTRNKRENSIFFLGLNTFTDFRHTINGPISPLMRSLTRSTVE

MLIHNWILTFSIFREHPSTVFQI**FTKCILVSSSFLLFYTLLPHGLLED**LMRRVGDSLVDLIVIC**EDSQGQHLSS****FCLFVATLQSPFSAGVSGLC**K
AILLPSKQIHVMIQSVDLSIGITNSLTNEQLCGFGFFLNVKTNLHCSRIPLITNLFLSARHMSLDLENSVGSYHPRMIWSVTWQWSNQVPAFGET
SLGFGMFQEKGQRHQNYEFPCRCIGTCGR**GSVQCAGLISLPIAIEFTY****QLTSSPTC**IVRPWRF**PNIFPLIACILLLSMNSTLSLFS**FSGGRSGY
VLMLSSKYQDSFTSKTRNKRENSIFFLGLNTFTDFRHTINGPISPLMRSLTRSTVE

Bilitranslokaza (QSAR)

Podatkovni nizi iz sodelovanja z Uni-TR

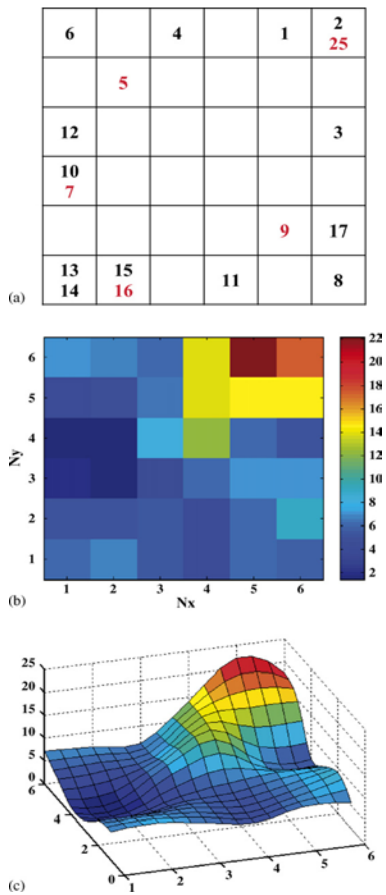
Strukturni deskriptorji: Codessa

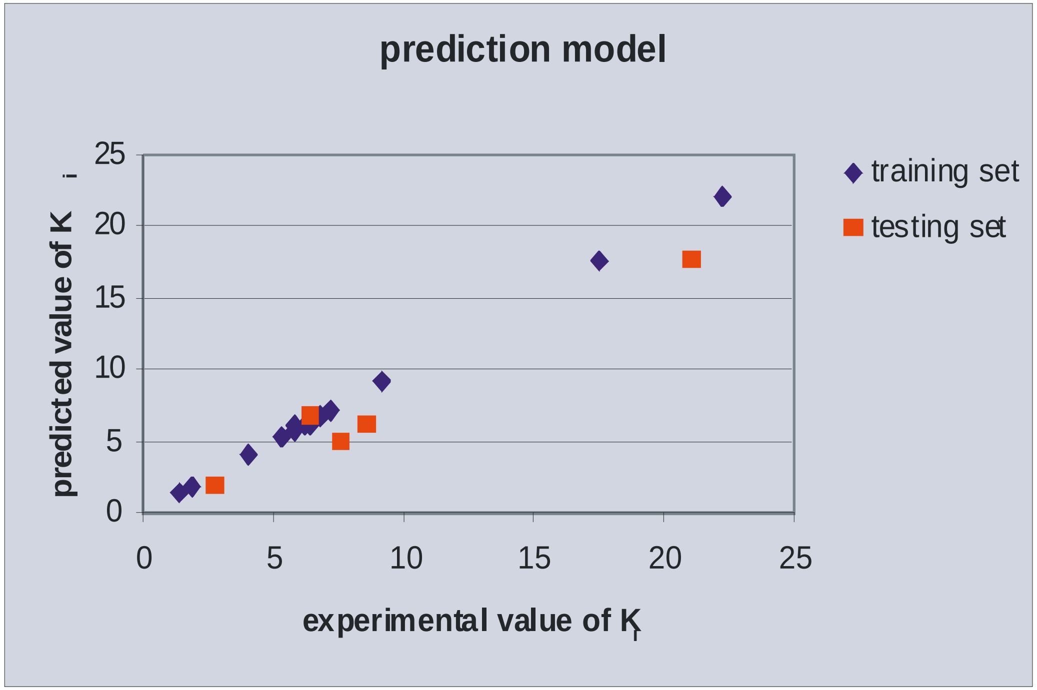
Lastnost: K_I (transport activity assay)

- Flavonoidi (flavonoli, heksociani, 20 + 20)
- Nukleotidi, nukleozidi (41)
- Naravne spojine, zdravilne učinkovine (37)

Bilitranslokaza (QSAR)

Napovedni model – BTL transport flavonoidov preko celične membrane



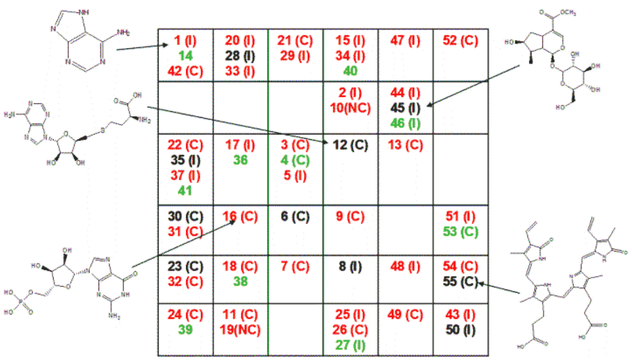


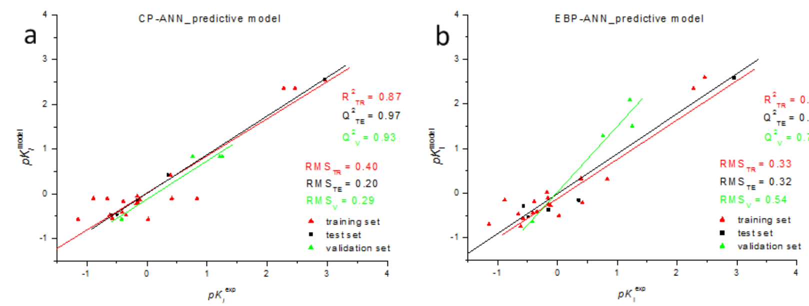
External test compound: sulfobromophthalein

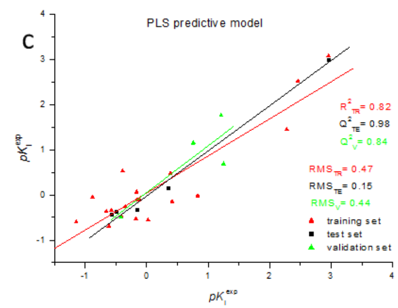
Ki (exp.) = 5.32 Ki (model) = 4.03

Karawajczyk A, Drgan V, Medic N, Oboh G, Passamonti S, Novič M. *Biochem. pharmacol.* 2007, **73**, 308-320. JCR IF: 3.581;
Župerl Š, Fornasaro S, Novič M, Passamonti S, *Anal. Chim. Acta* 2011, **705**, 322-333;
Roy Choudhury A, Župerl Š, Passamonti S, Novič M. *Acta Chim. Slov.* 2011, **58**, 385-392;

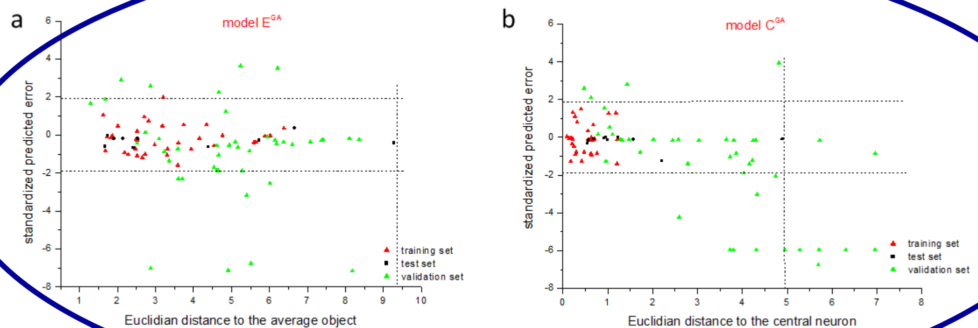
Napovedni model – BTL transport nukl.baz preko celične membrane







Domena uporabnosti



TR: 35

TE: 10

VA: 5+33

RMS_{VA} = 0.29

Karawajczyk A, Drgan V, Medic N, Oboh G, Passamonti S, Novič M. *Biochem. pharmacol.* 2007, **73**, 308-320. JCR IF: 3.581;
Župerl Š, Fornasaro S, Novič M, Passamonti S, *Anal. Chim. Acta* 2011, **705**, 322-333;
Roy Choudhury A, Župerl Š, Passamonti S, Novič M. *Acta Chim. Slov.* 2011, **58**, 385-392;

Bilitranslokaza 3D struktura?




MLIHNWILTFSIFREHPSTVFQI**FTKCILVSSSFLLFYTLLPHGLLED**LMRRVGDSLVDLIVIC**EDSQGQHLSSFCLFVATLQSPFSAGVSGLC**K
AILLPSKQIHVMIQSVDLSIGITNSLTNEQLCGFGFFLNVKTNLHCSRIPLITNLFLSARHMSLDLENSVGSYHPRMIWSVTWQWSNQVPAFGET
SLGFGMFQEKGQRHQNYEFPCRCIGTCGR**GSVQCAGLISLPIAIEFTYQLTSSPTC**IVRPWRF**PNIFPLIACILLLSMNSTLSLFS**FSGGRSGY
VLMLSSKYQDSFTSKTRNKRENSIFFLGLNTFTDFRHTINGPISPLMRSLTRSTVE

- Napoved TM domen (α-helix β-barrel)
- Eksperimentalna potrditev α-vijačnic
- ? Zanke AA znotraj in zunaj celice
- ? Monomera
- ? Transportni mehanizem

Napovedovanje transmembtranskih domen

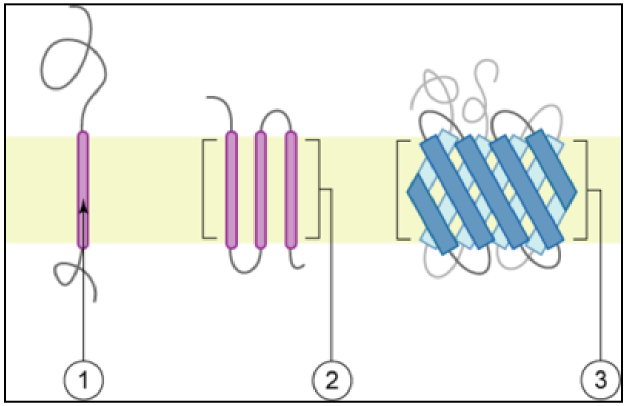
Segment: TWNIGVILLLTVMATAFMGYV

Amino Acid Adjacency Matrix

	A	C	G	I	L	M	F	P	W	V	R	N	D	E	Q	H	K	S	T	Y
A	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
I	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
M	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
V	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

ROWSUM : 2 0 2 2 3 2 1 0 1 2 0 1 0 0 0 0 0 0 3 1

nih TM proteinov iz PDB



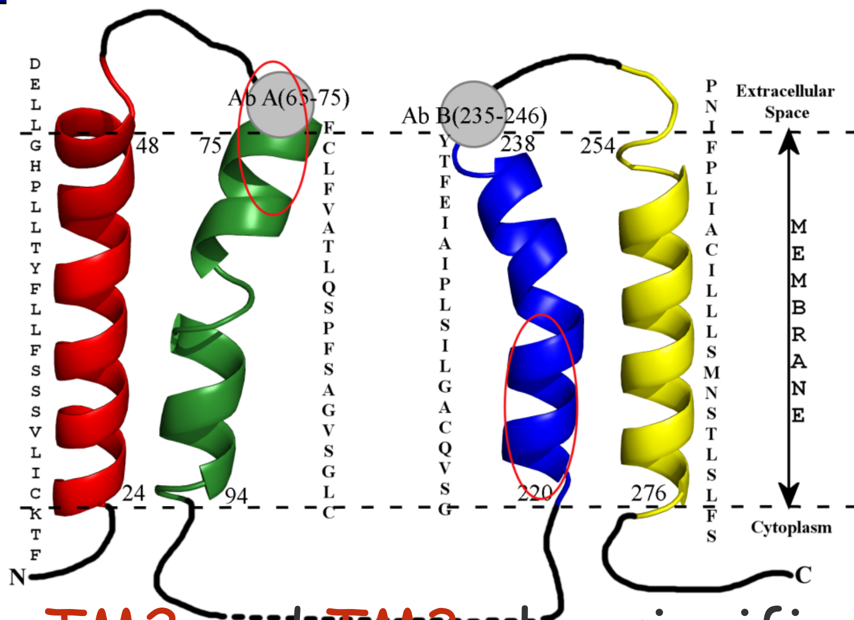
Pred α TM

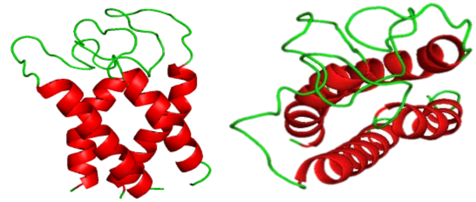
Pred β TM

http://www.ki.si/transmembrane-prediction/

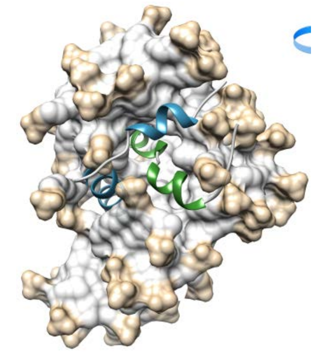
Roy Choudhury A, Novič M. *SAR &QSAR Environ. Res.* 2009, **20**, 741-754;
Randić M, Novič M, Roy Choudhury A, Plavšić D. *SAR &QSAR Environ. Res.* 2012, **23**, 327-343;
Roy Choudhury A, Novič M. *Int. J. Chem. Model.*, 2012, **4**, 205-219;
Roy Choudhury A, Zhukov N, Novič M. *The scientific world journal*, 2013, **2013**, 607830-1-607830-6

Bilitranslokaza





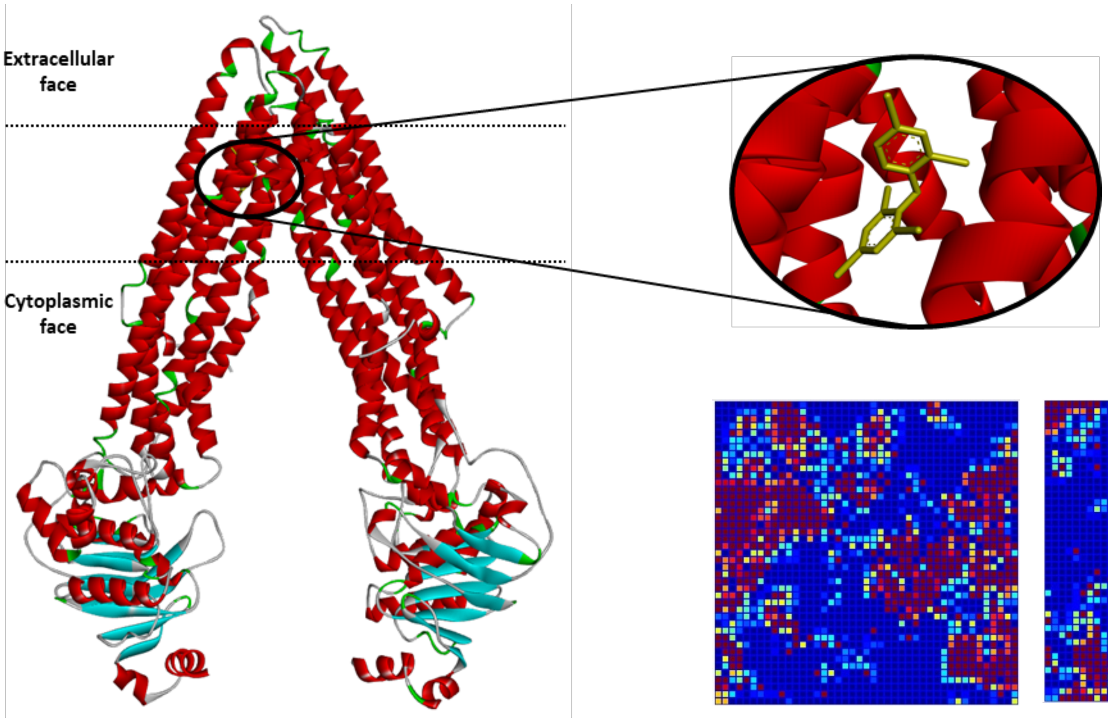
Z-DOPE=-0.55

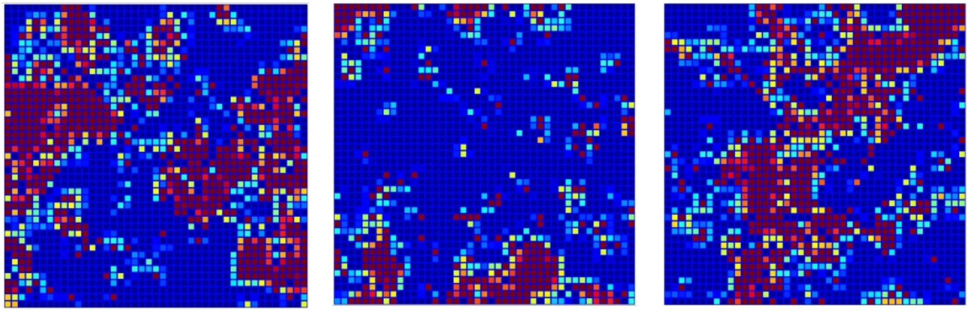


- TM2 and TM3 play significant role in transport channel formation, ligand binding and mediation
- Conserved Ser (73, 74, 229) and Cys (75, 224) are solvent accessible
- Role of H-bonding
- Probable allosteric nature
- Probable bi-directional transport system

Roy Choudhury A, Župerl Š, Sikorska E, Jurga S, Zhukov I, Novič M. *ACS Chemical Biology.* 2014, to be submitted
Roy Choudhury A, Perdih A, Župerl Š, Sikorska E, Šolmajer T, Jurga S, Zhukov I, Novič M. *BBA Biom.*2013, **1828**, 2609-2619;
Perdih A, Roy Choudhury A, Župerl Š, Sikorska E, Zhukov I, Šolmajer T, Novič M, *PloS one*, 2012, **7**, e38967-1-e38967-14;

P- Glycoprotein P-gp





P-gp: Inhibitors Substrates Inactive

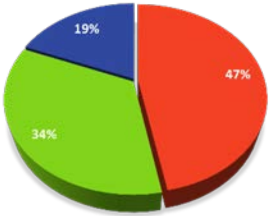
Liadys Mora Lagares, N. Minovski, M. Novič, Molecules 2019, 24, 2006; doi:10.3390/molecules24102006

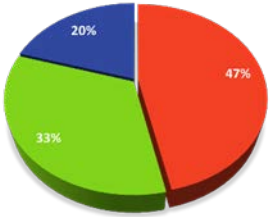
P- Glycoprotein P-gp

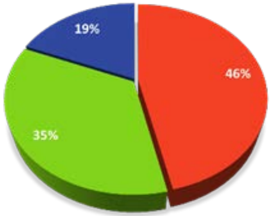
Training
1,786

Test
341

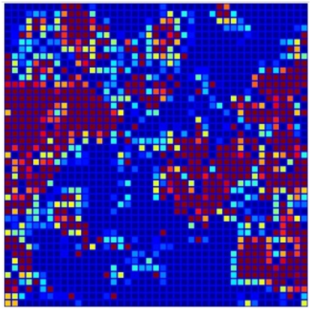
Validation
385

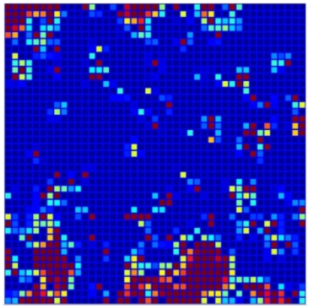


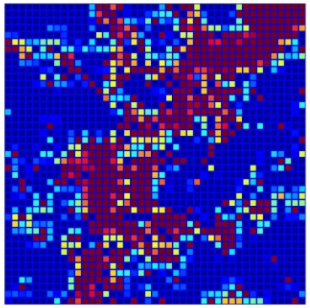




P-gp: Inhibitors Substrates Inactive

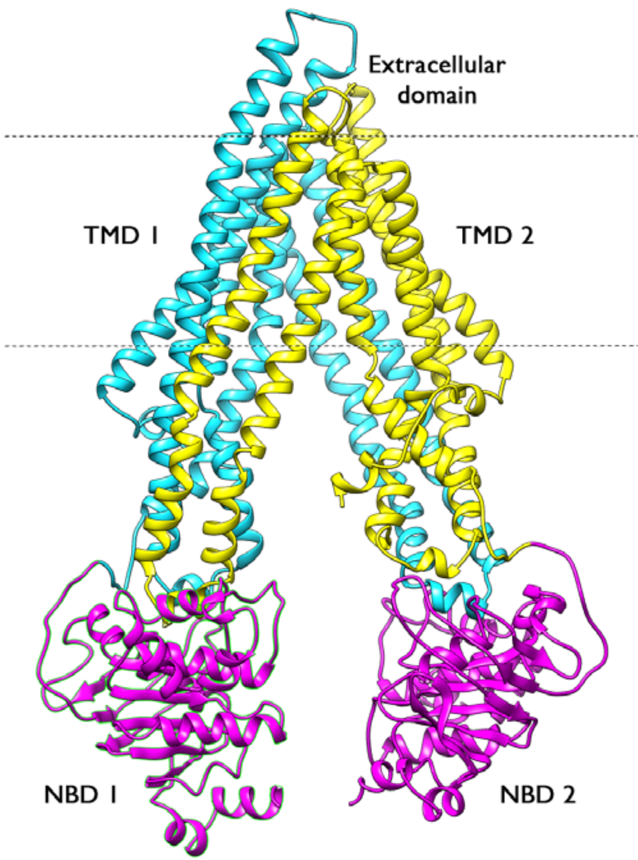




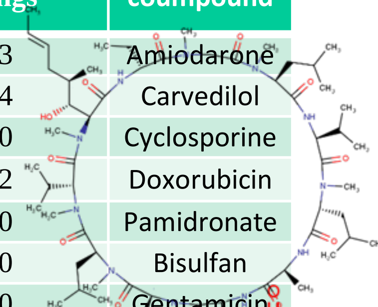


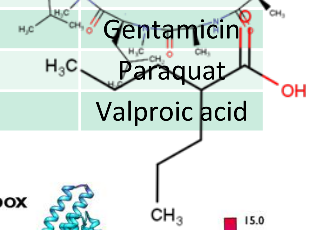
Liadys Mora Lagares, N. Minovski, M. Novič, Molecules 2019, 24, 2006; doi:10.3390/molecules24102006

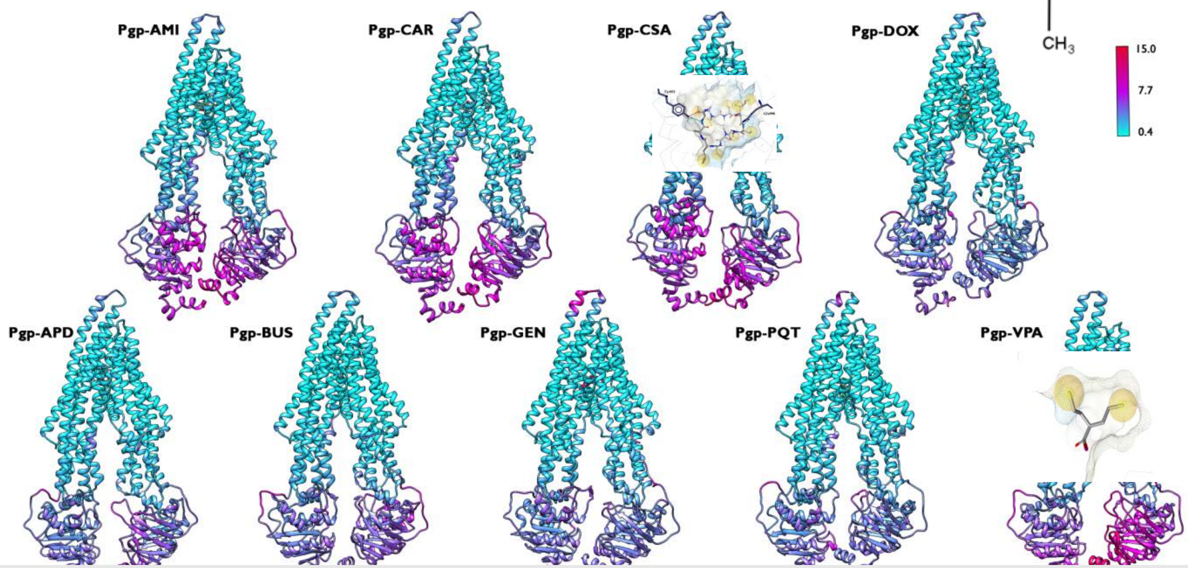
P- Glycoprotein P-gp



	LogP	HB D[j]	HB A[k]	TPSA[l] (Å²)	Heavy atom count	Aromatic rings	Name of the coumpound
AMI[a]	7.57	0	4	42.7	31	3	Amiodarone
CAR[b]	4.19	3	5	75.7	30	4	Carvedilol
CSA[c]	2.92	5	12	279.0	85	0	Cyclosporine
DOX[d]	1.27	6	12	206.0	39	2	Doxorubicin
APD[e]	-4.70	6	8	161.0	13	0	Pamidronate
BUS[f]	-0.52	0	6	104.0	14	0	Bisulfan
GEN[g]	-3.10	8	12	200.0	33	0	Gentamicin
PQT[h]	-4.22	0	0	7.8	14	2	Paraquat
VPA[i]	2.75	1	2	37.3	10	0	Valproic acid



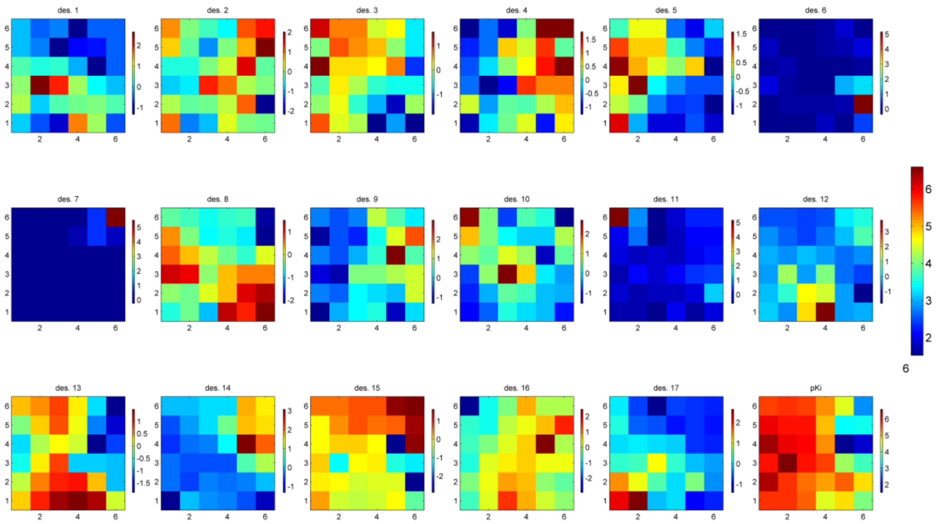


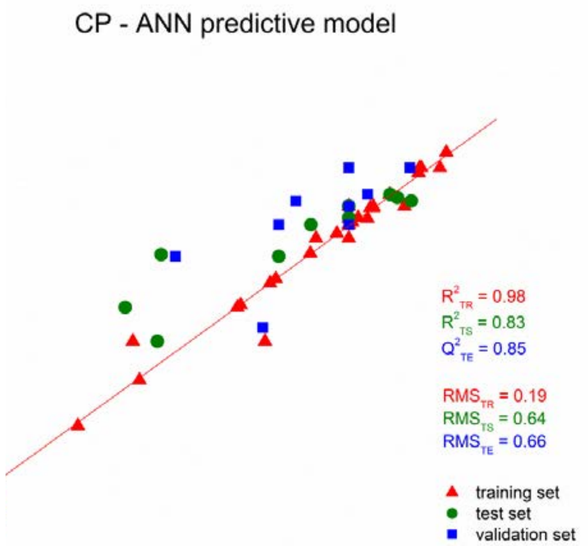


Liadys Mora Lagares, N.Minovski, A.Y.Caballero Alfonso, E.Benfenati, S.Wellens, M.Culot, F.Gosselet, M. Novič, Int. J. Mol. Sci. 2020, 21, 4058; doi:10.3390/ijms21114058

Liadys Mora Lagares et al., Structure-function relationships in ABCB1: insights from molecular dynamics simulations, Sent for publication

Katepsin K

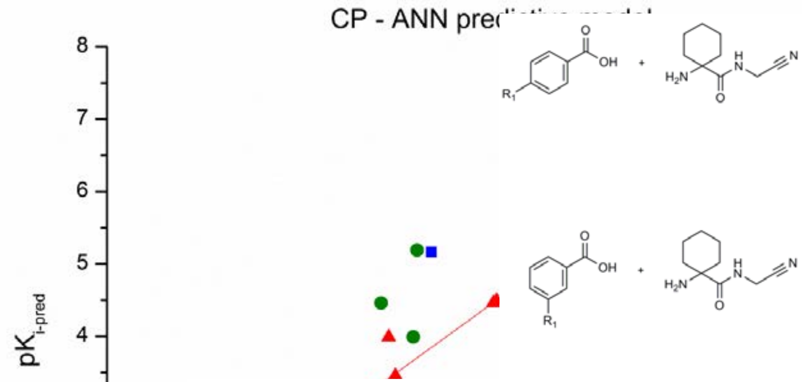


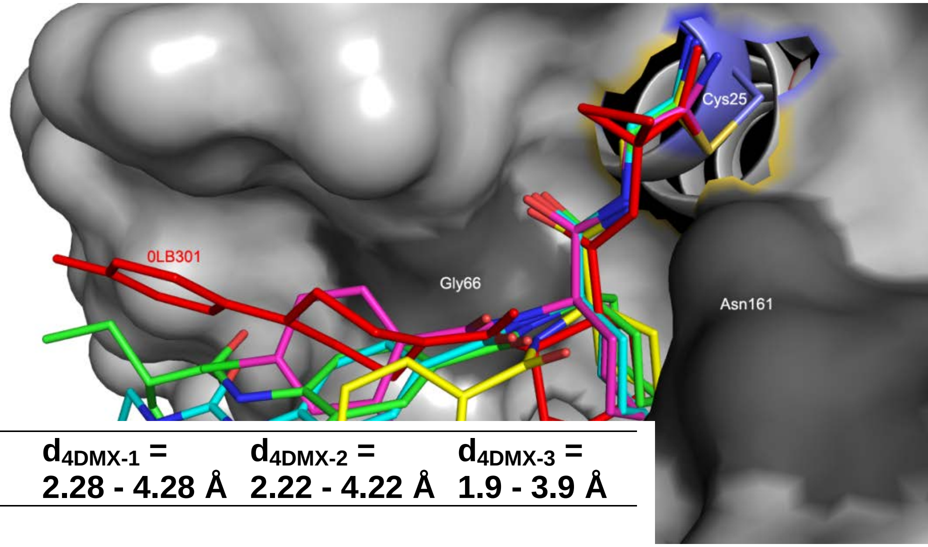


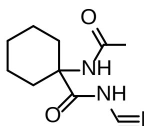
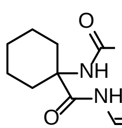
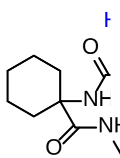
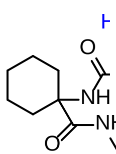
Descriptors associated with the covalent binding are selected because of the nitrile warhead that binds covalently to the cysteine residue of the enzyme in the S1 binding site.

On the other hand, the descriptors associated with molecular shape were identified as important, implying the accommodation of the P2 and P3 moieties of the inhibitors in the S2 and S3 binding sites of the enzyme and contributing to the inhibitor–enzyme interactions.

Katepsin K





ID	Chemical structure	CSF	$K_{i\text{-pred}}$ (nM)	$pK_{i\text{-pred}}$	$d_{4DMX\text{-}1}$ = 2.28 - 4.28 Å	$d_{4DMX\text{-}2}$ = 2.22 - 4.22 Å	$d_{4DMX\text{-}3}$ = 1.9 - 3.9 Å
1	 R1	35.0	66.1	4.18	3.25	2.91	2.80
2	 R2	33.3	1.40	5.85	3.25	2.76	2.95
3	 R3	32.4	6.85	5.16	3.23	2.75	2.98
4	 R4	33.6	66.1	4.18	3.29	2.83	2.92

Hvala za pozornost