EVOLUTION OF INHIBITION MODELS FOR FLUORINATED *CINCHONA* ALKALOIDS BY MACHINE LEARNING

Ana Mikelić, mag. chem.

Department of Chemistry, University of Zagreb Faculty of Science, Horvatovac 102a, Zagreb, Croatia

ana.mikelic@chem.pmf.hr

This research was founded by the Croatian Science Foundation, grant numbers IP-2016-06-3775 „ADESIRE" and ESF-DOK-2018-09-3416

Computational Chemistry Day 2022

Zagreb, 24. 9. 2022.

Content

Background and Objective

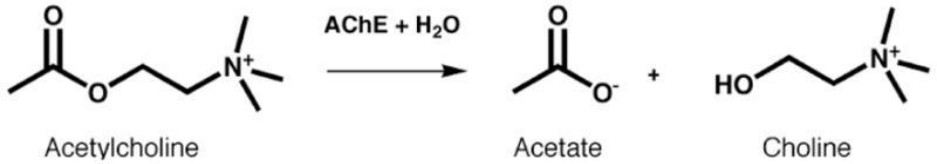
- **Alzheimer's disease (AD)** ⇒ one of the most common forms of dementia – chronic syndrome of the CNS
  - *Multifactorial disease → difficulties with treatment*
  - *The most successful approach to date: improving cholinergic transmission using **cholinesterase inhibitors***
- **Cholinesterases (ChEs):** acetylcholinesterase (AChE) and butyrylcholinesterase (BChE) → hydrolysis of acetylcholine (ACh)
- Normal conditions: ACh is dominantly decomposed by **AChE**; the physiological role of **BChE** is still unclear
- Progressed AD: level of **AChE** in brain = 55-67% of normal values – level of **BChE** = 120% of normal values

  → **Specific BChE** or **dual inhibitors** (BChE and AChE) might be a potential therapeutic strategy to be utilized in the treatment of AD
- **Derivatives of *Cinchona* alkaloids** ⇒ proved to be potent BChE inhibitors with high selectivity toward BChE

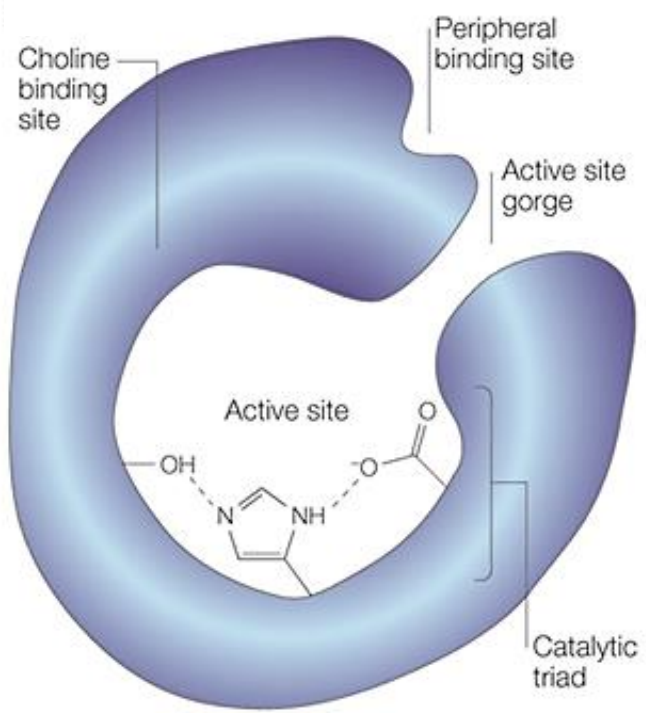
1. K. Sharma, *Mol. Med. Rep.* **20** (2019) 1479–1487.
2. Q. Li, H. Yang, Y. Chen, H. Sun, *Eur. J. Med. Chem.* **132** (2017) 294-309.
3. A. Bosak, A. Ramić, T. Šmidlehner, T. Hrenar, I. Primožič, Z. Kovarik, *PLoS One* **13** (2018) 1–18.

Cholinesterases

- Enzyme-catalyzed reaction:



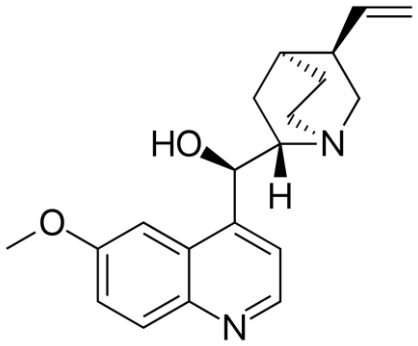
- Structure:
  - *AChE and BChE share **65%** amino acid sequence homology*
  - *<u>Catalytic triad</u>: Ser, His,Glu*
- Turnover number: $\sim 1.5 \times 10^4\ s^{-1}$



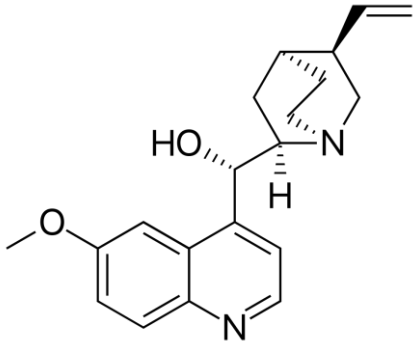
4. H. Soreq and S. Seidman, *Nat. Rev. Neusrosci.* **2** (2001) 294–302.
5. P. Sharma, P. Srivastava, A. Seth, P. N. Tripathi, A. G. Banerjee, S. K. Shrivastava, *Prog. Neurobiol.* **174** (2018) 53–89.
6. P. Taylor, S. Camp, Z. Radić, *Acetylcholinesterase* in L.R. Squire (Ed.), *Encyclopedia of Neuroscience*, Academic Press, Oxford (2009), pp. 5-7.

Cinchona alkaloids

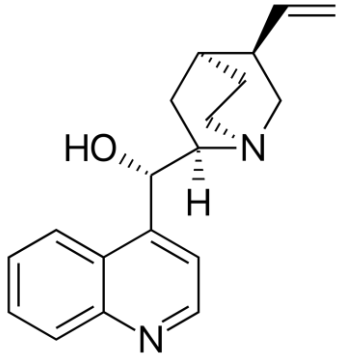
- ***Cinchona* alkaloids** ⇒ natural products isolated from the bark of the *Cinchona* tree
- The most researched: quinine (**Q**), quinidine (**QD**), cinchonine (**CN**), and cinchonidine (**CD**)
- Bioactivity



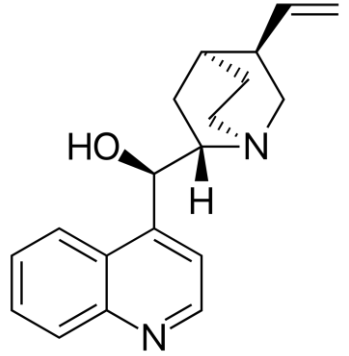
(Q)



(QD)



(CN)



(CD)

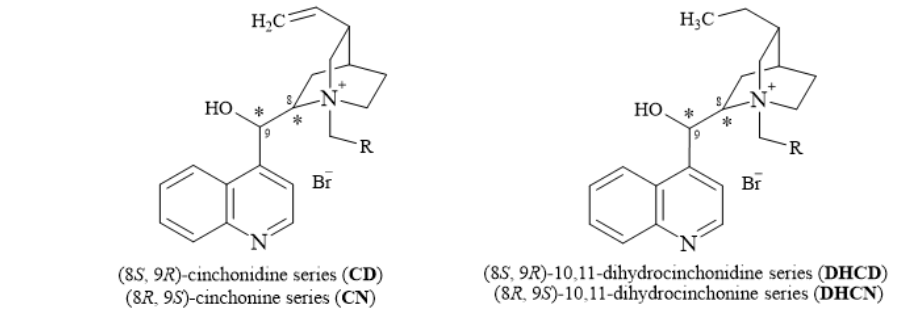
7. P. J. Boratyński, M. Zielińska-Błajet, J. Skarżewski, *Alkaloids Chem. Biol.* **82** (2019) 29–145.

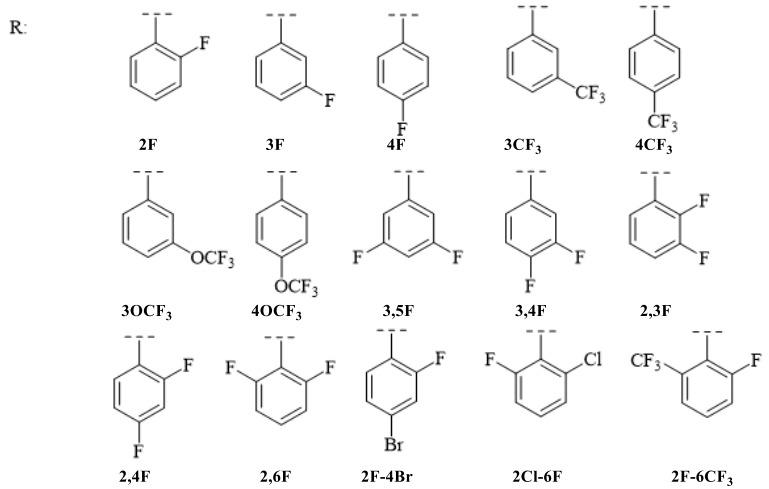
Methodology

- Creating an activity/PES model using machine learning multivariate linear regression

Investigated compunds

- A series of **46** *Cinchona* alkaloid derivatives which differ in positions of fluorine atom(s) in the molecule were previously prepared:





Activity data

- Dissociation constants of the enzyme-inhibitor complex (K_i) $\Rightarrow$ measure of compound's inhibition potency
- **Principal components** of the activity data extracted by 2nd-order tensor decomposition using our own program ***moonee***

Compound	K_i/μM		SI*	Compound	K_i/μM		SI*
	BChE	AChE			BChE	AChE	
CD Bzl	0.075±0.007	15±2		CN Bzl	2.9±0.3	121±12	
CD 2F	0.82 ± 0.03	33 ± 1	40	CN 2F	2.4 ± 0.1	80 ± 2	33
CD 3F	0.075 ± 0.005	40 ± 2	533	CN 3F	6.1 ± 0.3	13 ± 0.4	2.1
CD 4F	1.5 ± 0.1	69 ± 3	46	CN 4F	2.6 ± 0.1	3.9 ± 0.2	1.5
CD $3CF_3$	2.4 ± 0.1	34 ± 1	14	CN $3CF_3$	4.4 ± 0.2	59 ± 2	13
CD $4CF_3$	2.0 ± 0.1	21 ± 1	11	CN $4CF_3$	6.0 ± 0.3	31 ±1	5.2
CD 3,5F	0.081±0.01	10±1	123	CN 3,5F	6.3 ± 0.2	34 ± 3	5.4
CD 3,4F	1.3 ± 0.1	13 ± 1	10	CN 3,4F	6.1 ± 0.2	14 ± 0.2	2.3
CD 2,3F	0.75 ±0.03	19 ±1	25	CN 2,3F	9.6 ±0.4	46 ±3	4.8
CD 2,4F	6.1 ± 0.5	6.4 ± 0.3	1.1	CN 2,4F	6.0 ±0.2	27 ±1	4.5
CD 2,6F	9.9 ±0.4	7.7 ± 0.5	0.77	CN 2,6F	5.2 ±0.2	30 ±2	5.8
CD $3OCF_3$	7.4 ±0.4	8.2 ± 1.1	1.1	CN $3OCF_3$	4.7 ± 0.2	41 ±2	8.7
CD $4OCF_3$	7.6 ±0.5	7.3 ±0.5	0.96	CN $4OCF_3$	7.8 ±0.3	19 ± 2	2.4
CD 2F-$6CF_3$	5.7 ±0.6	35 ±4	6.1	CN 2F-$6CF_3$	7.7 ±0.4	61 ±2	7.9
CD 2F-4Br	0.68 ±0.05	7.2 ±0.4	10	CN 2F-4Br	5.5 ± 0.3	16 ±1	2.9
CD 2Cl-6F	5.0 ±0.3	9.9 ± 0.8	1.9	CN 2Cl-6F	1.2 ±0.0	17 ±1	14
DHCD	19±2	206±6	11	DHCN	1.2±0.1	43±2	43
DHCD Bzl	0.4±0.02	4.8±0.4	12	DHCN Bzl	0.9x±0.04	21±1	23
DHCD 3F	0.3±0.02	27±2	84	DHCN 3F	1.2±0.1	20±1	20
DHCD 4F	4.3±0.2	31±1	8	DHCN 4F	1.6±0.1	64±2	40
DHCD $3CF_3$	1.4±0.1	25±1	18	DHCN $3CF_3$	1.2±0.05	41±2	34
DHCD $4CF_3$	3.2±0.2	15±1	5	DHCN $4CF_3$	1.6±0.1	18±1	9
DHCD $3OCF_3$	6.8±0.3	36±1	5	DHCN $3OCF_3$	1.3±0.5	68±1	52
DHCD $4OCF_3$	5.9±0.2	17±1	3	DHCN $4OCF_3$	2.2±0.1	22±1	10

* SI denotes stereoselective index calculated from K_i(AChE)/K_i(BChE)

8. T. Hrenar, ***moonee***, Program for Manipulation and Analysis of Multi-and Univariate Data, rev. 0.68268, 2022.

Potential energy surfaces (PES)

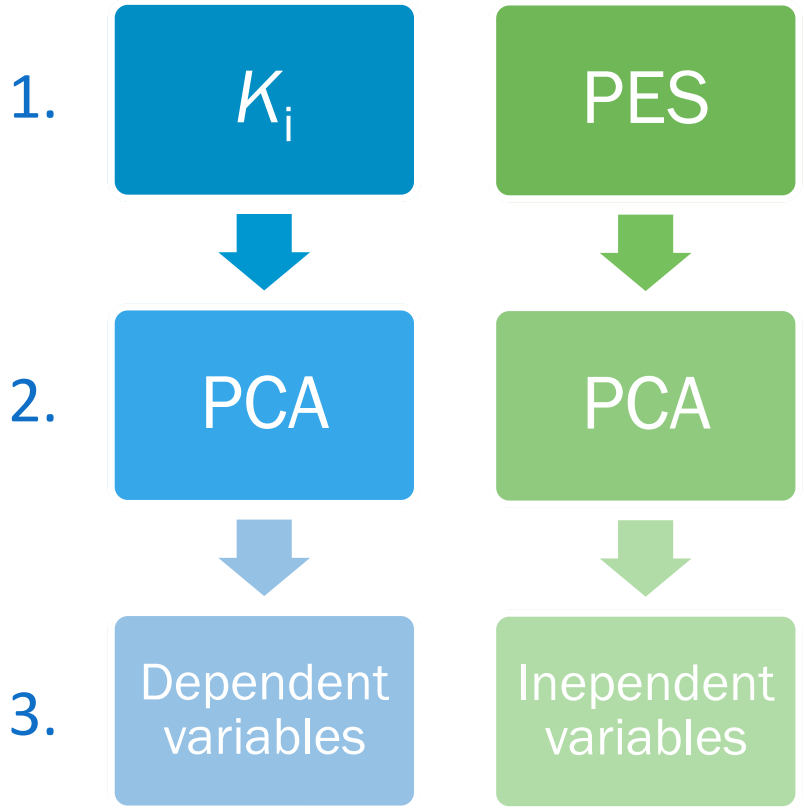
- **Potential energy surfaces** (**PES**) were sampled by performing *ab initio* molecular dynamics simulations using our own program ***qcc***:
  - On-the-fly *calculations of forces in each point of the simulation using the semi-empirical PM7 Hamiltonian implemented in MOPAC2016*
  - *Integration: velocity Verlet algorithm with a 0.5 fs step size*
  - *Initial temperature: 773.15 K; kept constant using a velocity scaling algorithm*
  - *Total length of the simulation: 2.5 ps (total of 5 000 000 steps for each compound)*
- Construction of **strict local maxima plateaus** by counting all strict local maxima points in the probability distribution funcions $\Rightarrow$ criterion for conformational space coverage
- Multidimensional PES were reduced in dimension by principal component analysis

9. T. Hrenar, ***qcc***, Quantum Chemistry Code, rev. 0.6826, 2022.
10. J. J. P. Stewart, *Stewart Computational Chemistry*; MOPAC2016: Colorado Springs, CO, USA, 2016.

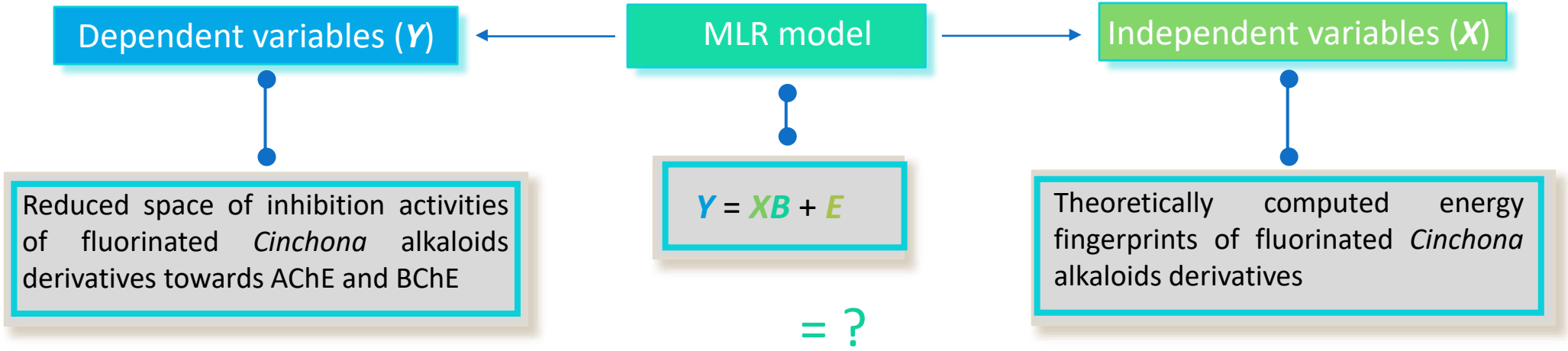
Generating principal components

- **Principal components analysis** (**PCA**): eigenvalues and eigenvectors of the covariance matrix $\boldsymbol{C_X} = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}^{\tau}$

1. Data (***K***$_i$ or **PES** of **CD**, **CN**, **DHCD** and **DHCN**, and 46 of their derivatives – total of 50 compounds) were:
   - exported to the ASCII format,
   - *arranged* in the matrix ***X*** with numbers written in a free format,
   - *mean-centered*
2. PCA on the covariance matrix was carried out using our own multivariate analysis code ***moonee*** based on the NIPALS algorithm
3. The most important principal components were used as dependent or independent variables in multivariate linear regression



Multivariate linear regression (MLR) model



$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \cdot \begin{bmatrix} b_{01} & b_{02} & \cdots & b_{0p} \\ b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mp} \end{bmatrix} + \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{np} \end{bmatrix}$$

n – number of samples

m – number of independent variables

p – number of dependent variables

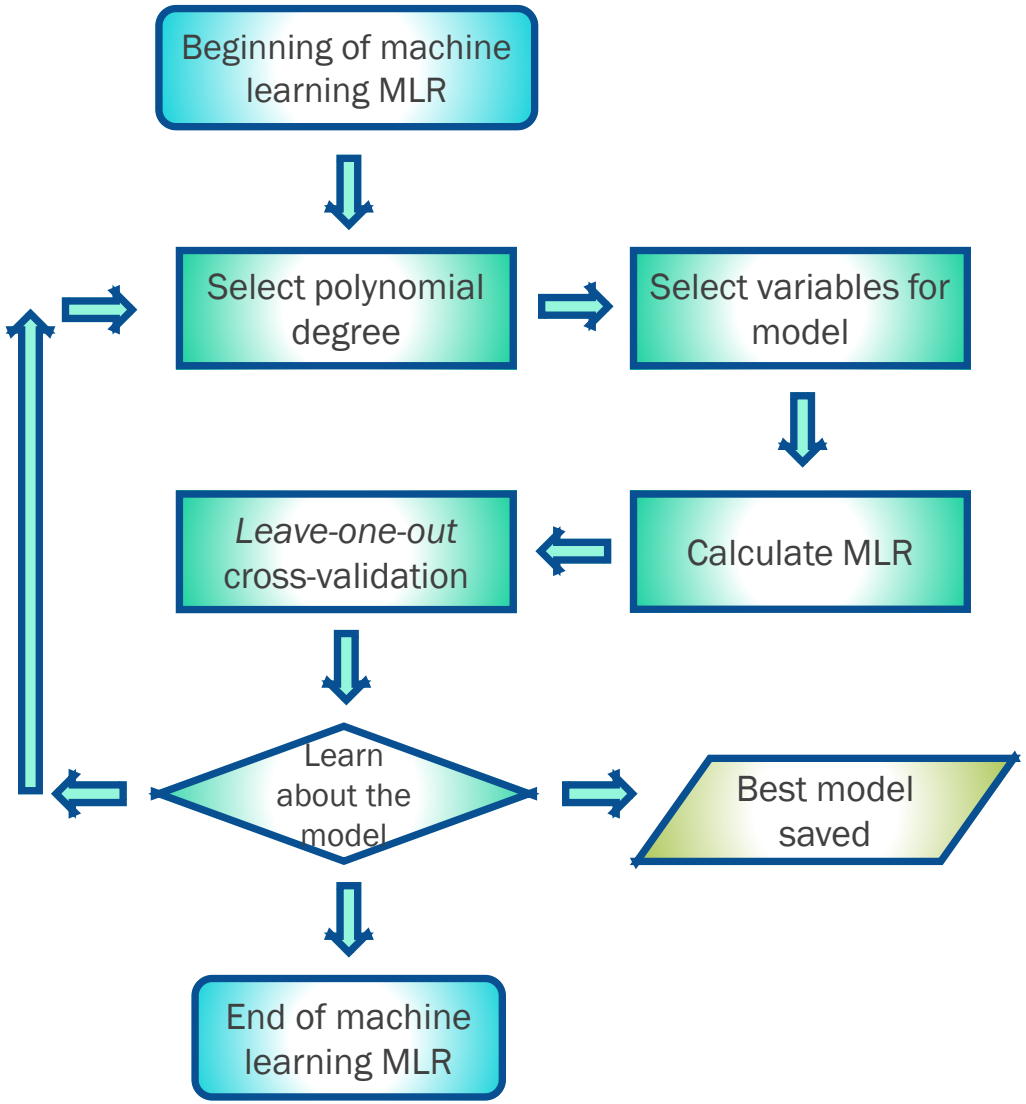
Machine learning multivariate linear regression (ML-MLR)

- **Multivariate linear regression** was performed using the following expression for matrices of coefficients $\boldsymbol{B}$ calculated by singular value decomposition:

$$\boldsymbol{B} = (\boldsymbol{X}^{\tau}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{Y}$$

- An **extensive machine learning procedure** was applied for generating multivariate linear regression models with linear combination of original variables as well as their higher-order polynomial terms
- Models were thoroughly validated by the ***leave-one-out* cross-validation** technique (LOO-CV)

Machine learning (ML)



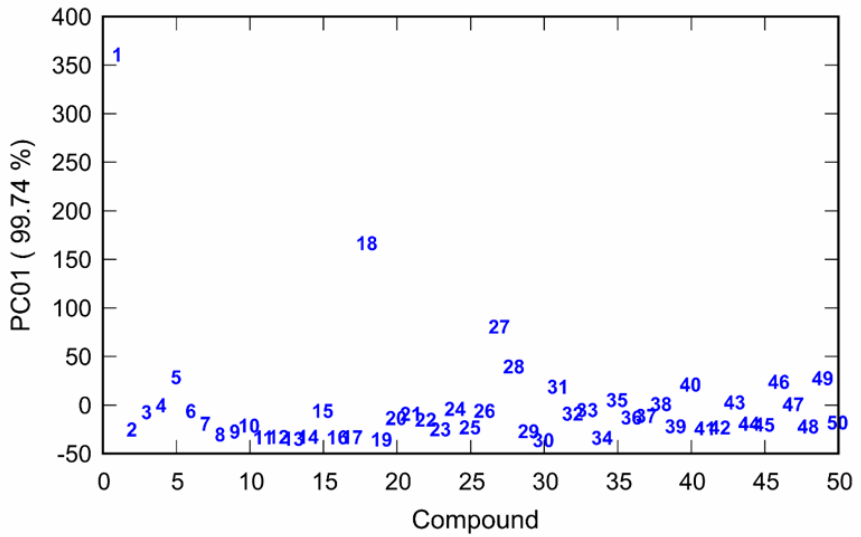
Results

Classification model based on activities

- PCA on the mean-centered covariance matrix of *Cinchona* alkaloids derivatives inhibition of AChE and BChE showed that the **first** principal component explained more than **99.74%** of the total variance among the data

  → The 1st component is the most important in describing the inhibition activity of the compounds

- Possibility of **visualization**

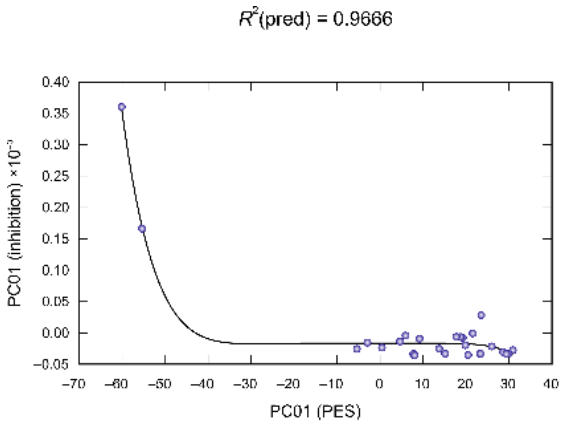


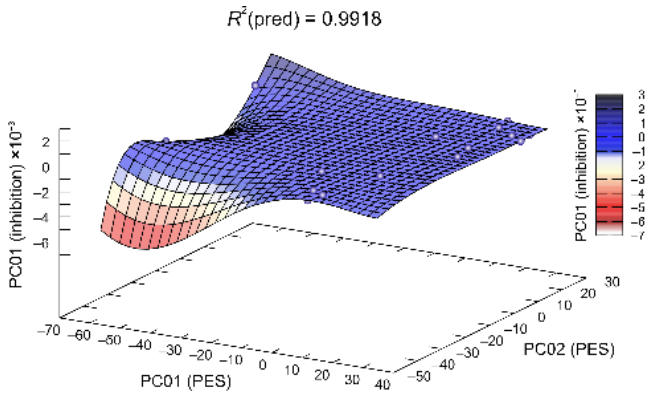
Activity/PES models

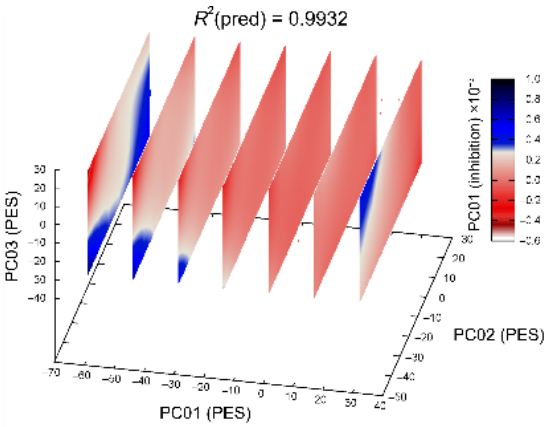
- Models were created using the the **first three** principal components of the reduced PES data and the **1st** principal component for inhibition of AChE and BChE
- All possible regression models were generated, and the ***B***-matrices of coefficients were determined
- **Machine learning** was used to determine the best possible regression model
- **3D models** were inspected up to the fourth order, and the total number of investigated models in each case was 17 179 869 184

Evolution of regression models

- Optimal activity/PES models were selected based on the adjusted and predicted R^2 values, LOO-CV mean squared error, as well as the number of variables in the models
- Progression of the best inhibition models for fluorinated *Cinchona* alkaloids derivatives depending on domain dimensionality:



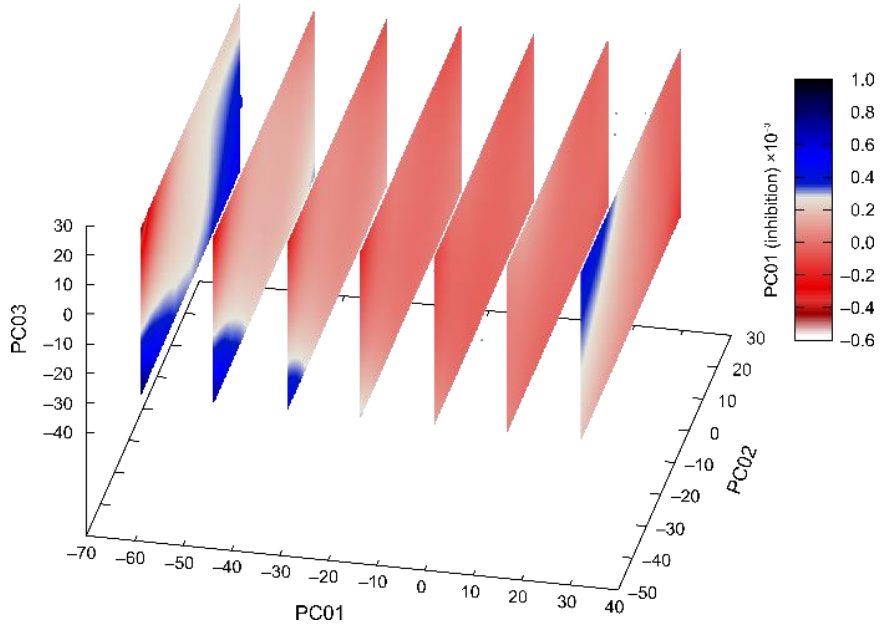




The best regression model

- The 1st principal component for flourinated **CD**-derivatives inhibition of AChE and BChE was regressed on compounds' theoretically computed energy fingerprints, and the best calculated 3D regression model was determined by machine learning:

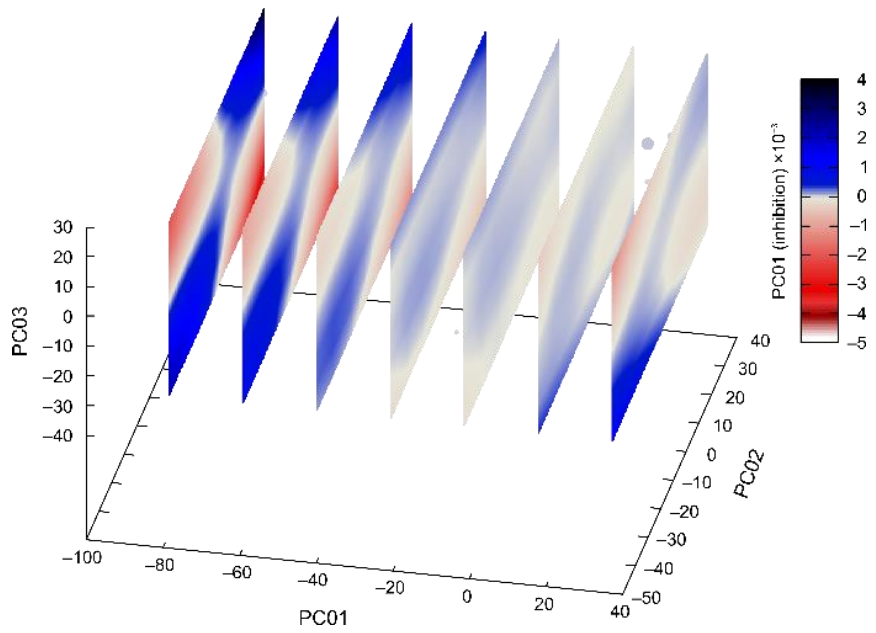
$$R^2_{\text{adj}} = 0.9936 \qquad R^2_{\text{pred}} = 0.9932$$



The best regression model

- The 1st principal component for flourinated **CN**-derivatives inhibition of AChE and BChE was regressed on compounds' theoretically computed energy fingerprints, and the best calculated 3D regression model was determined by machine learning:

$R^2_{\text{adj}} = 0.9849$ $R^2_{\text{pred}} = 0.9879$



Conclusion

- Activity and PES data were reduced in dimension by a 2nd order tensor decomposition tool, PCA.
- The best multivariate linear regression models describing the relationship between activity and PES were determined by machine learning.
- The best activity/PES models can be (and were) used for prediction of inhibition activities for new compounds based only on their theoretically computed PES.
- *N*-quaternary derivatives of *Cinchona* alkaloids proved to be an excellent scaffold for further research towards finding selective cholinesterase inhibitors.

Thank you!