Using artificial neural networks in molecular modelling

Dalibor Trapl

4.4.2019

2nd meeting of PhD_post.doc club

Department of Biochemistry and Microbiology
UCT PRAGUE

https://biomikro.vscht.cz/

Supervisor:
doc. Ing. Vojtěch Spiwok, PhD.

Outline

- What are artificial neural networks?
- How can we simply describe complex atomistic motion

  -> introduction to collective variables
- Examples of complex collective variable
- How can we express them in form of artificial neural network
- Use of these collective variables (CVs)

  -> metadynamics
- Plans for future

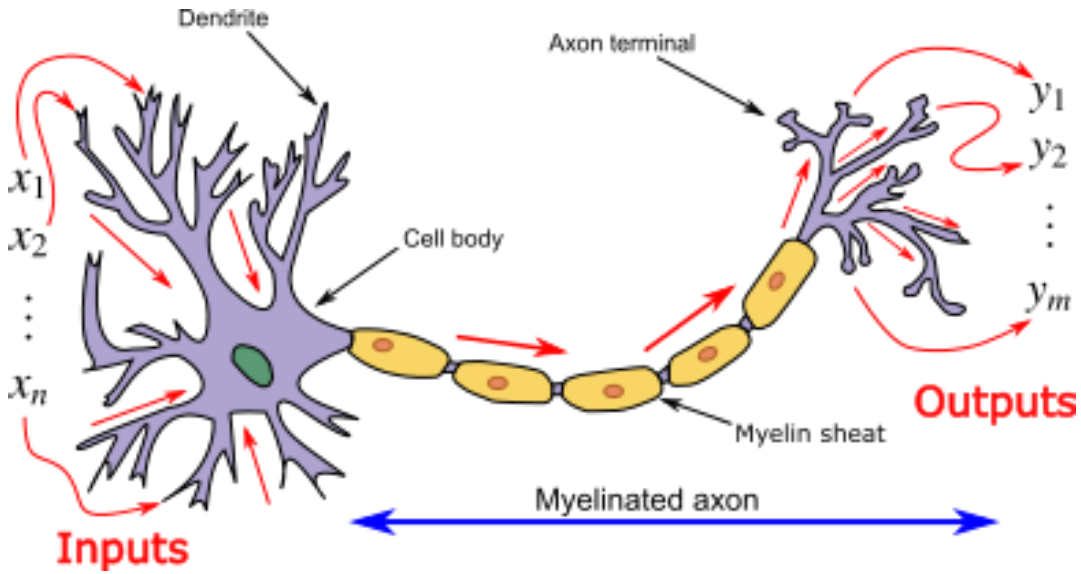
Let's start with a non-scientific example

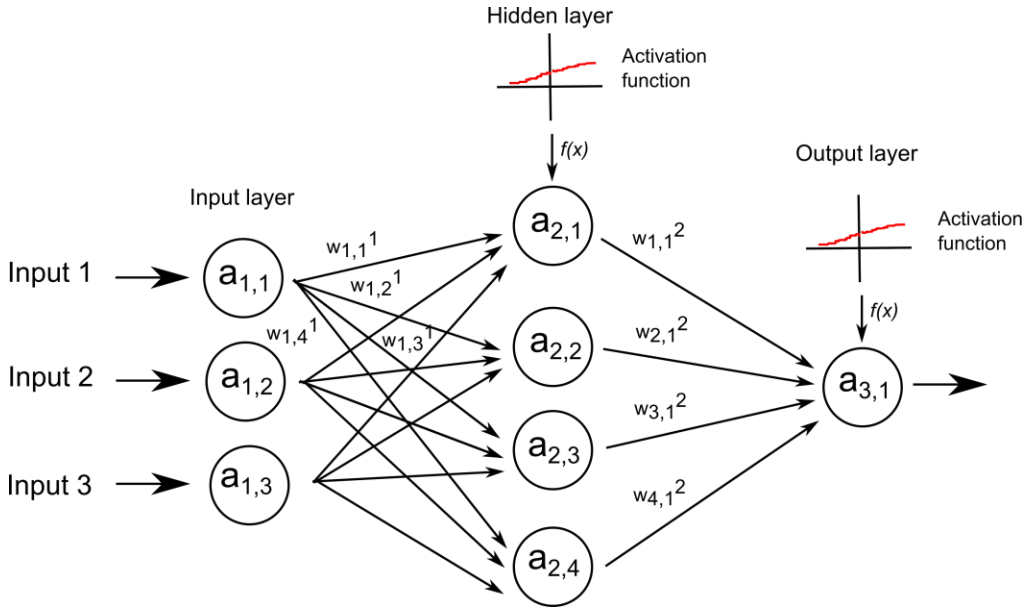
- "Mushroom-picking" - national sport in Czechia

- How to recognize different mushrooms – task for ANN

Artificial neural network (ANN)

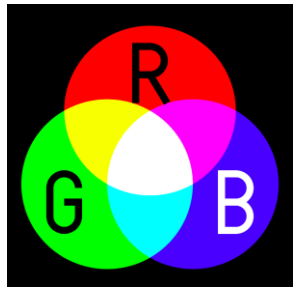
- Term of computer science (AI) known since 1950s
- Based on real biological neurons





https://en.wikipedia.org/wiki/Artificial_neural_network

ANN – mushroom recognizing app



N neurons = 3x number of pixels

0...1

0...1

0...1

0...1

0...1

Hidden layer(s)

N neurons = Number of mushroom in database

90 %

https://play.google.com/store/apps/details?id=bazinac.aplikacenahouby&hl=cs

Aplikace na houby

Vocom Vzdělávání

4 026

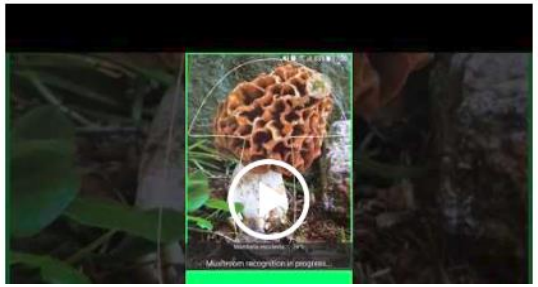
PEGI 3

Nabídka nákupů v aplikaci

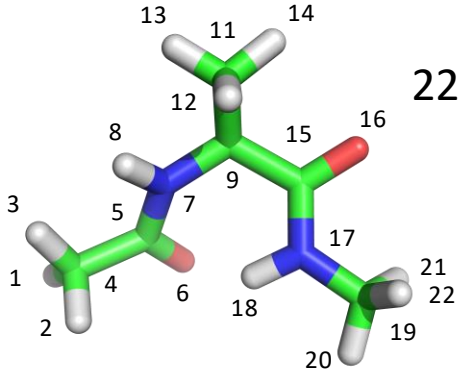
Tato aplikace je kompatibilní s vaším zařízením.

Přidat do seznamu přání

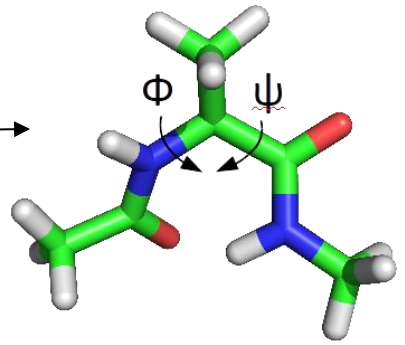
Nainstalovat



Back to science - Collective variables (CVs)



22 atoms in 3D space
=
66 coordinates



2 dihedral angles
=>
Dimensionality reduction
without significant loss
of information

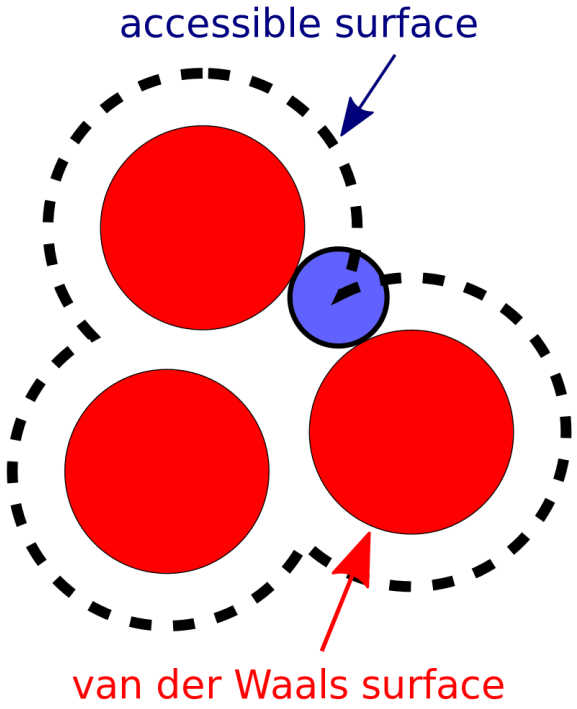
There are a lot of different CVs:

distance, angle, dihedral angle, radius of gyration, alphaRMSD, ...

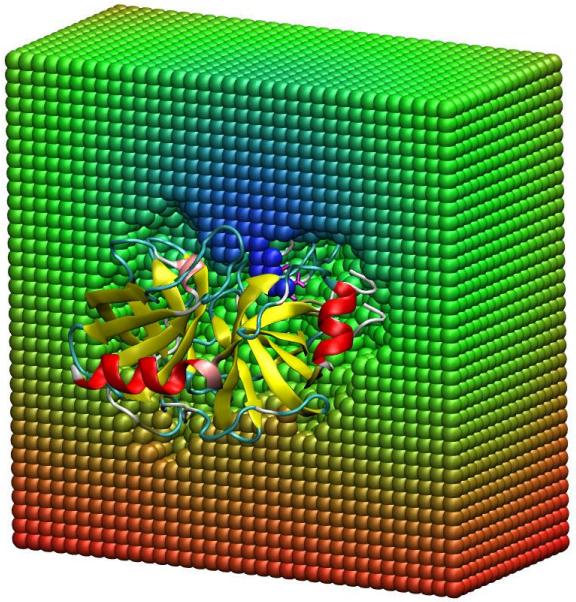
Complexity

Example of complex CVs

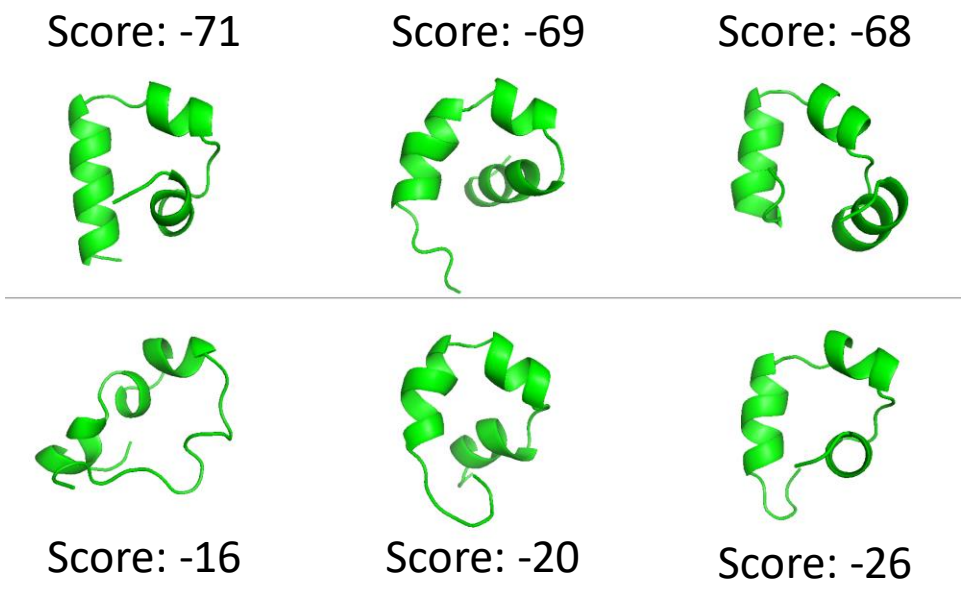
Solvent accessible surface area (SASA)



Distance field distances

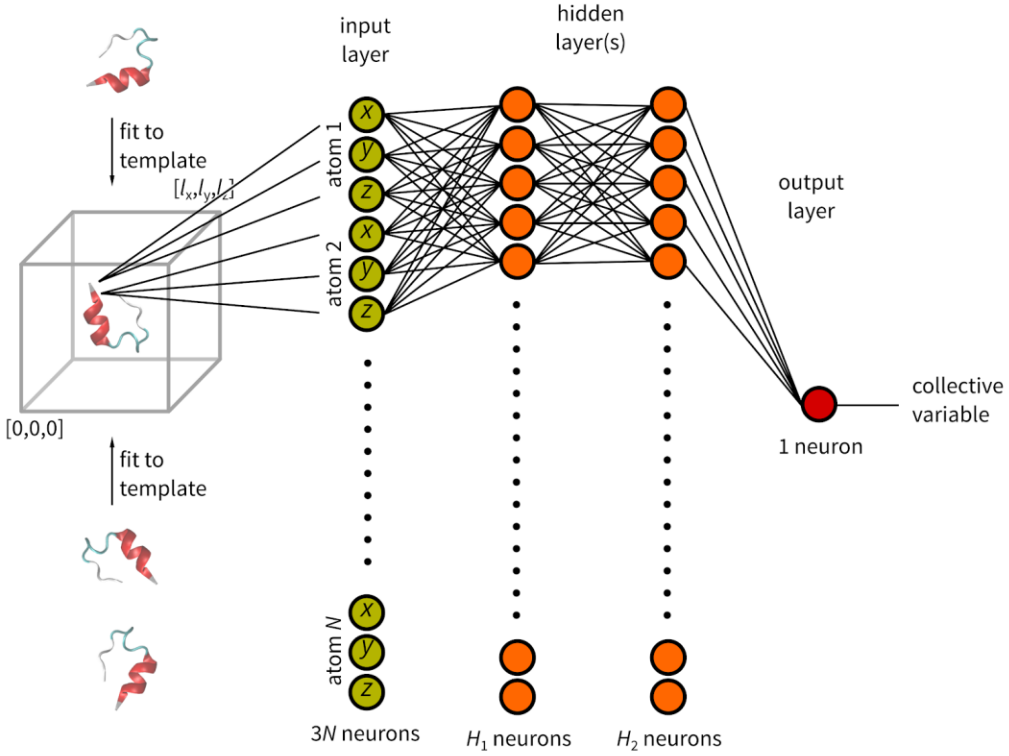


Rossetta scoring function
for protein models



ANNCOLVAR

- Available on https://github.com/spiwokv/anncolvar



Plumed input file


```
WHOLEMOLECULES ENTITY0=1-582
FIT_TO_TEMPLATE STRIDE=1 REFERENCE=ref.pdb TYPE=OPTIMAL
p1: POSITION ATOM=1 NOPBC
          .
          .
p582: POSITION ATOM=582 NOPBC
p1x: COMBINE ARG=p1.x COEFFICIENTS=0.166667 PERIODIC=NO
p1y: COMBINE ARG=p1.y COEFFICIENTS=0.166667 PERIODIC=NO
p1z: COMBINE ARG=p1.z COEFFICIENTS=0.166667 PERIODIC=NO
          .
          .
p582x: COMBINE ARG=p582.x COEFFICIENTS=0.166667 PERIODIC=NO
p582y: COMBINE ARG=p582.y COEFFICIENTS=0.166667 PERIODIC=NO
p582z: COMBINE ARG=p582.z COEFFICIENTS=0.166667 PERIODIC=NO
l1_1: COMBINE ARG=p1x,p1y,p1z,...,p581z,p582x,p582y,p582z COEFFICIENTS=-0.022379,...,0.020054
PERIODIC=NO
          .
          .
l1_64: COMBINE ARG=p1x,p1y,p1z,...,p582x,p582y,p582z COEFFICIENTS=-0.027287,...,-0.097164 PERIODIC=NO
l1r_1: MATHEVAL ARG=l1_1 FUNC=1.0/(1.0+exp(-x-0.025188)) PERIODIC=NO
          .
          .
l1r_64: MATHEVAL ARG=l1_64 FUNC=1.0/(1.0+exp(-x+0.034216)) PERIODIC=NO
l2: COMBINE ARG=l1r_1,...,l1r_64 COEFFICIENTS=-2.333472,...,0.064548 PERIODIC=NO
l2r: MATHEVAL ARG=l2 FUNC=(x-2.248804) PERIODIC=NO
PRINT ARG=l2r STRIDE=100 FILE=COLVAR
```

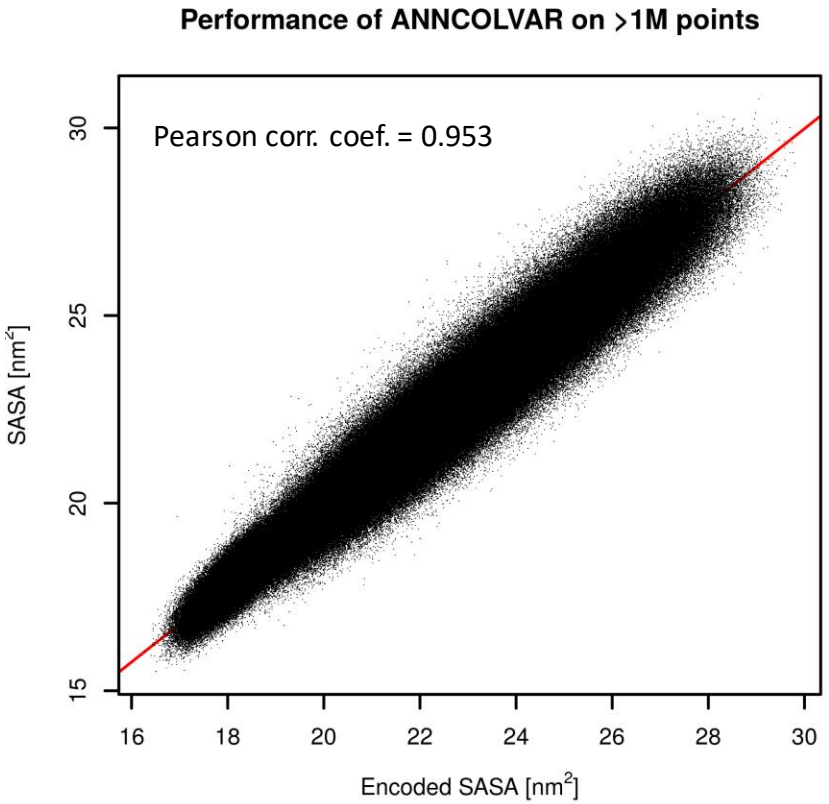
Performance of ANNCOLVAR

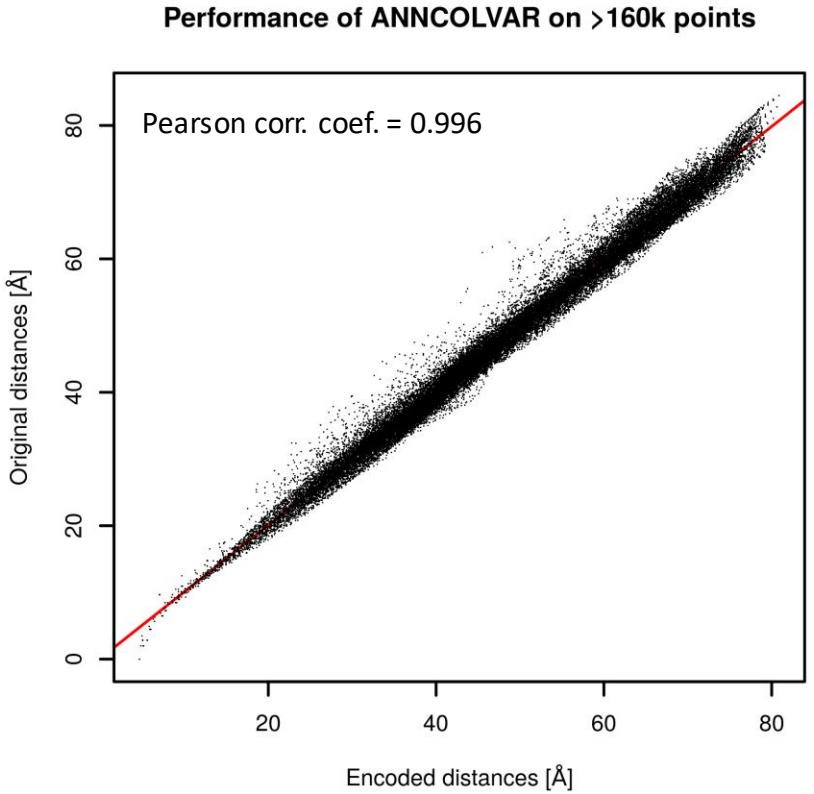
Rosetta Commons
The hub for Rosetta modeling software

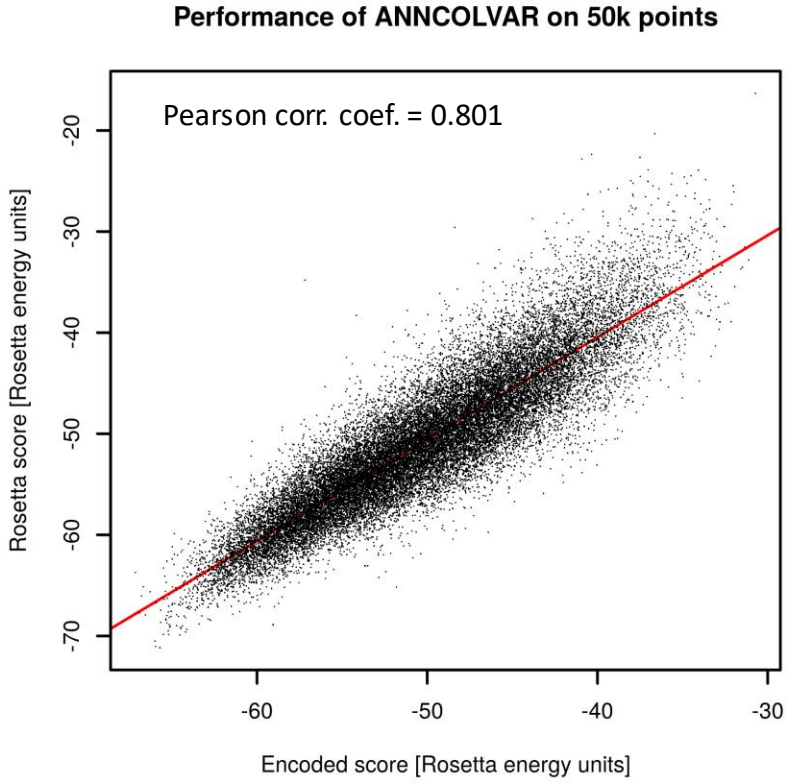
SASA of trpcage calculated 208-μs trajectory of Trp-cage
Lindorff-Larsen, K., Piana, S., Dror, R.O., Shaw, D.E. (2011)
How fast-folding proteins fold. *Science*. 334(6055), 517–520.
doi: 10.1126/science.1208351

Distance field distances calculated by Dijkstra's algorithm
Complex trypsin-benzamidine
https://github.com/spiwokv/distancefield/distancefield.c

50k models of chicken villin headpice
generated by Rosetta Abinitio
(de novo structure prediction)







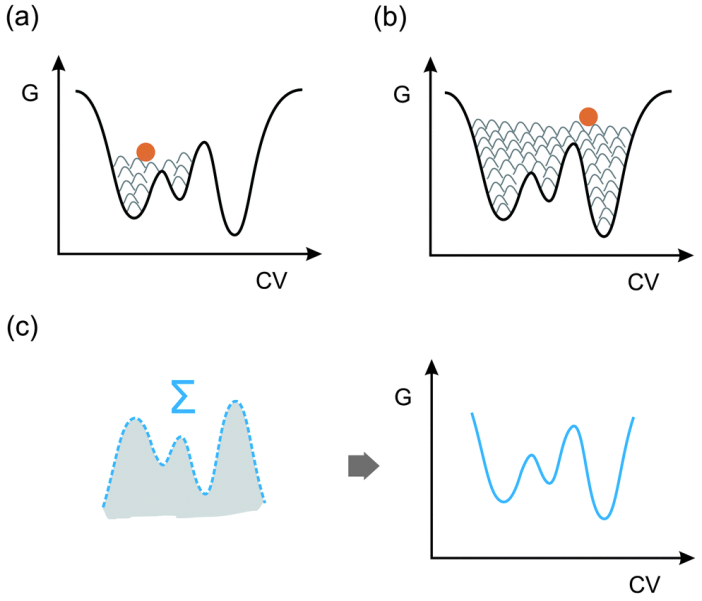
Trapl, D., Horvaćanin, I., Mareška, V., Özçelik, F., Spiwok, V., Unal, G. (2019)
anncolvar: Approximation of Complex Collective Variables by Artificial Neural Networks for Analysis and Biasing of Molecular Simulations.
Front. Mol. Biosci. **accepted**
doi: 10.3389/fmolb.2019.00025

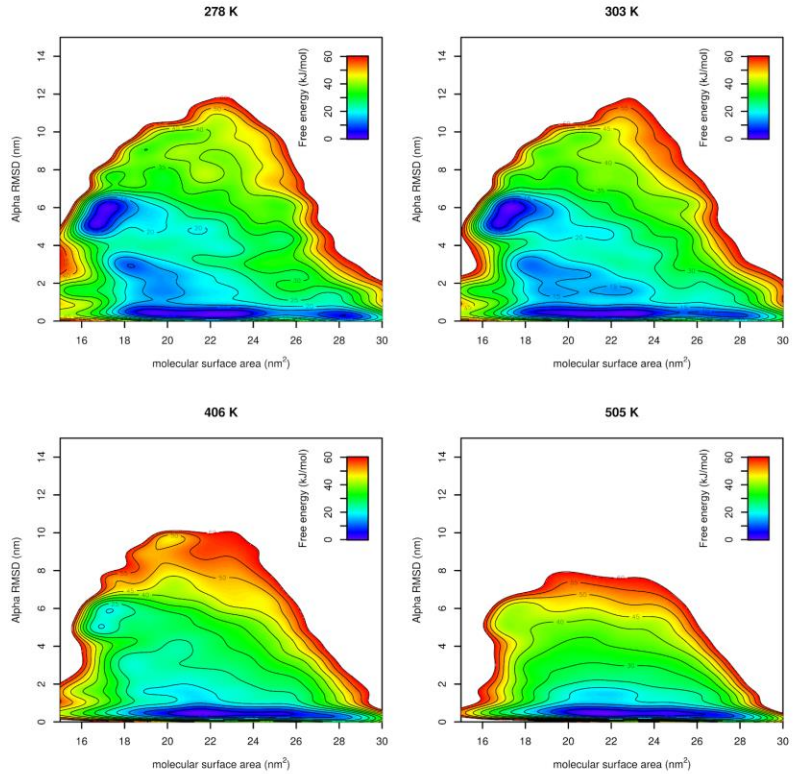
Bc. Thesis (defense in June 2019) :
Protein-ligand interaction studied by a combination of metadynamics and distance field
Lucie Hrdá

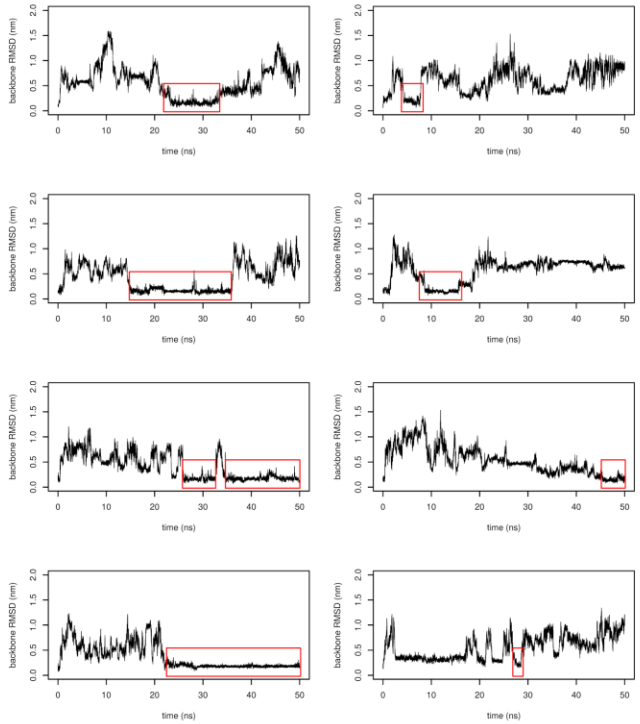
Master Thesis (defense in June 2020) :
Sampling enhancement in molecular simulations aided by *de novo* protein structure prediction techniques
Kateřiina Tomášková

Metadynamics using SASA approximated by ANN

Movie recommendation:
https://youtu.be/SOq2xiYsGbM







Van Speybroeck, V. *et al.* First principle chemical kinetics in zeolites: the methanol-to-olefin process as a case study. *Chem. Soc. Rev.* **43**, 7326–7357 (2014).
doi: 10.1039/c4cs00146j. Epub

Trapl, D., Horvaćanin, I., Mareška, V., Özçelik, F., Spiwok, V., Unal, G. (2019)
anncolvar: Approximation of Complex Collective Variables by Artificial Neural Networks for Analysis and Biasing of Molecular Simulations. *Front. Mol. Biosci.* **accepted**
doi: 10.3389/fmolb.2019.00025

METADYNAMICA

http://www.metadynamics.cz/
https://web.vscht.cz/~spiwokv/research.html

Happy end :-)

ORIGINAL RESEARCH ARTICLE Provisionally accepted The full-text will be published soon. Notify me

Front. Mol. Biosci. | doi: 10.3389/fmolb.2019.00025

anncolvar: Approximation of Complex Collective Variables by Artificial Neural Networks for Analysis and Biasing of Molecular Simulations

Dalibor Trapl[1], Izabela Horvaćanin[1, 2], Vaclav Mareška[1], Furkan Özçelik[3], Vojtech Spiwok[1*] and Gozde Unal[3]

[1]University of Chemistry and Technology in Prague, Czechia

[2]Faculty of Science, University of Zagreb, Croatia

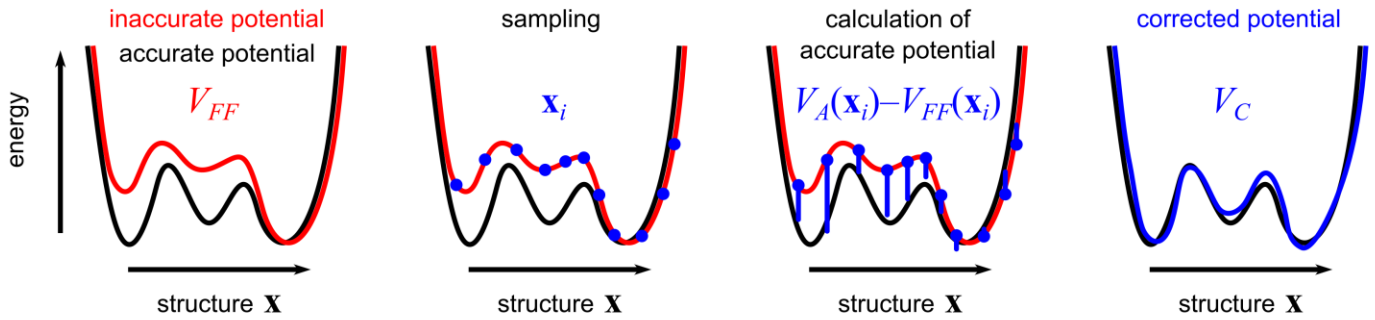
[3]Department of Computer Engineering, Faculty of Computer and Informatics Engineering, Istanbul Technical University, Turkey

The state of a molecular system can be described in terms of collective variables. These low-dimensional descriptors of molecular structure can be used to monitor the state of the simulation, to calculate free energy profiles or to accelerate rare events by a bias potential or a bias force. Frequent calculation of some complex collective variables may slow down the simulation or analysis of trajectories. Moreover, many collective variables cannot be explicitly calculated for newly sampled structures. In order to address this problem, we developed a new package called anncolvar. This package makes it possible to build and train an artificial neural network model that approximates a collective variable. It can be used to generate an input for the open-source enhanced sampling simulation PLUMED package, so the collective variable can be monitored and biased by methods available in this program. The computational efficiency and the accuracy of anncolvar are demonstrated on selected molecular systems (cyclooctane derivative, Trp-cage miniprotein) and selected collective variables (Isomap, molecular surface area).

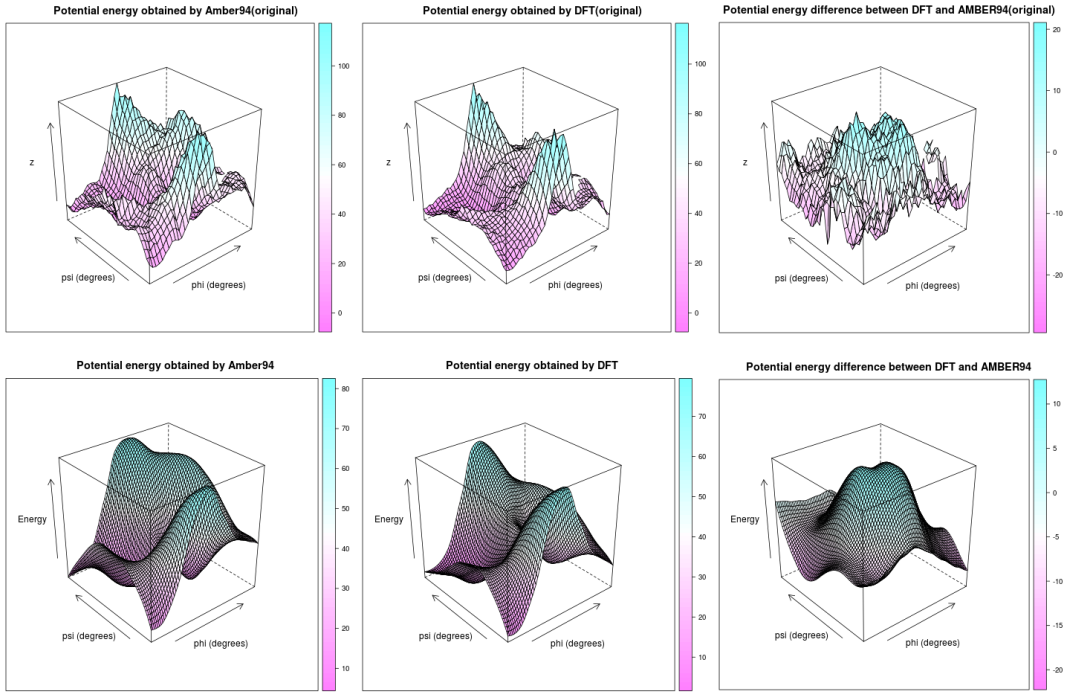
Keywords: Metadynamics, neural networks, molecular dynamics simulation, Collective variables, Free energy simulations

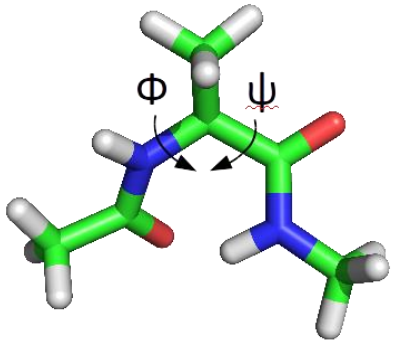
Received: 22 Jan 2019; **Accepted:** 01 Apr 2019.

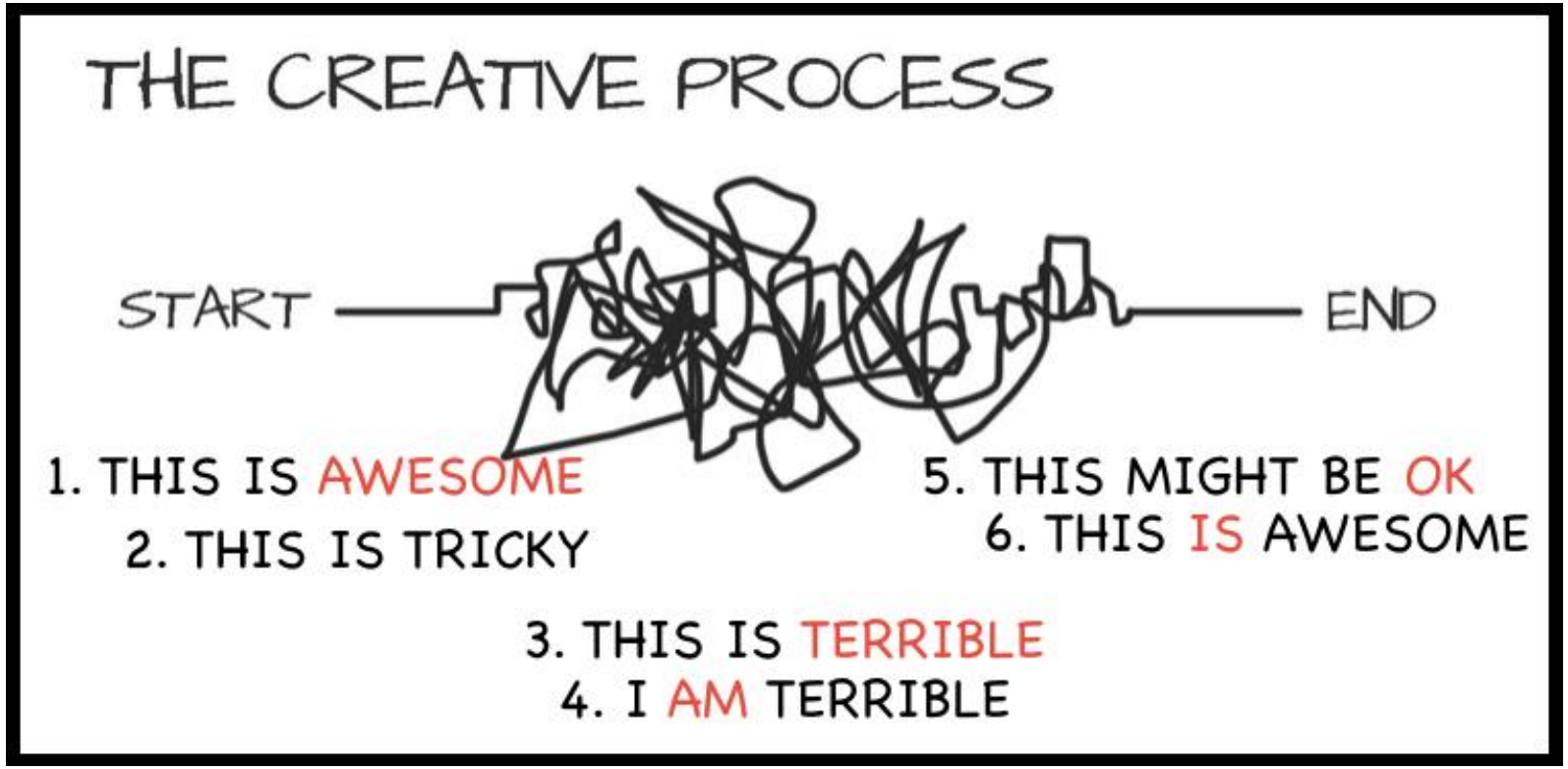
Plans for future



Correction of molecular mechanics force fields







Founding

elixir CZECH REPUBLIC

GAČR CZECH SCIENCE FOUNDATION

Financial support from specific university research (MSMT No 21-SVV/2019)

Computational resources

cesnet metacentrum

Many thanks to:

Vojtěch Spiwok

Lucie Hrdá

Kateřina Tomášková

University of Zagreb/ UCT intern

Izabela Horvaćanin

UCT Prague

Václav Mareška

Istanbul Technical University

Furkan Özçelik

Gozde Unal