Combing molecular dynamics and machine learning for advanced pose and free-energy prediction

Oleksandr Yakovenko (Alex), PhD
Chief Technical Officer
IFOWON.CO (a spinoff from BC Cancer Agency)
Vancouver , Canada
http://www.ifowonco.com/

Overview

1. Pose prediction with molecular dynamics, docking and neuron networks (ranked ~2)

2. Free energy evaluations with combination of Jarzynski pulling and umbrella sampling (ranked ~1)

Motivation

for further interest in high-throughput virtual screening

One of our goals at IFOWON.CO is offer services for identification of compounds that do not inhibit but **reprogram** proteins i.e. modify their functionality and binding preferences.

For example, creating a synthetic pathway in a living cell by gluing two otherwise non-interacting proteins, yields so large number of therapeutic possibilities that it is becoming economically difficult to physically screen all opportunities, even with large automated screening platforms.

In addition there are no high-throughput screening assays available and even worse, such screens can be high-risk investments because traditional confirming knockdown assays of RNAi, that we can use for a typical inhibitor assay are not applicable for development of reprogramming drugs.

Therefore we utilize an extensive up-front computational assessment to identify reprograming compounds. This means that only a few dozen compounds need to be tested with expensive and labor-intensive generic (pulldown) assays and majority of cost is put into cheap and flexible theoretical research.

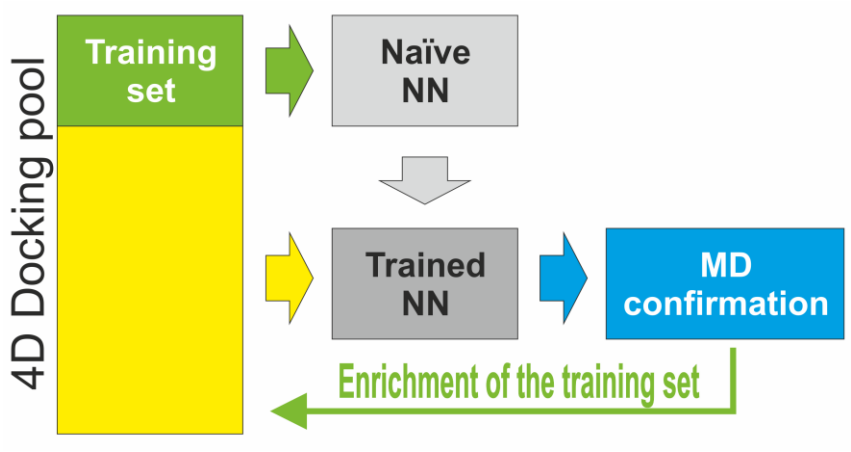
Two real-life examples of modulating protein interactions that are under development at BC Cancer Agency:

- perturbing interpretation a of epigenetic mark with a compound that reprogram a mark reading protein so that it changes own histone mark binding preferences

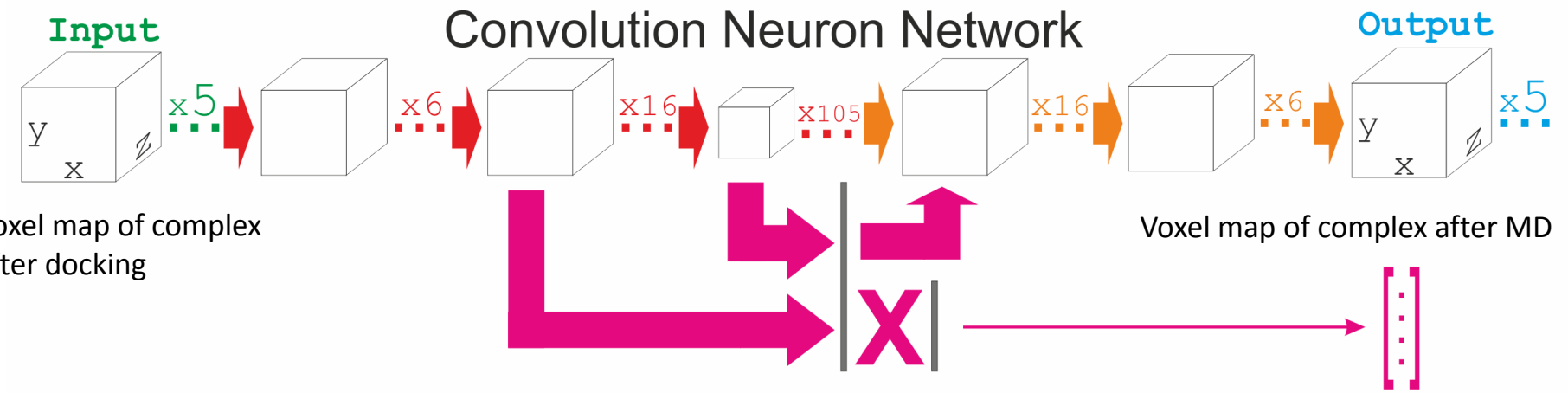
- a modulating compound, which “hijacks” an ubiquitin ligase, to induce degradation of the target protein

In both cases I had carried out intensive *in silico* research and then several selected compounds were tested in a wet lab with slow and labor-intensive generic pulldown assay to end up with the hits.

Integration of convolution neuron networks with molecular dynamics (MD) and docking for accurate pose predictions.



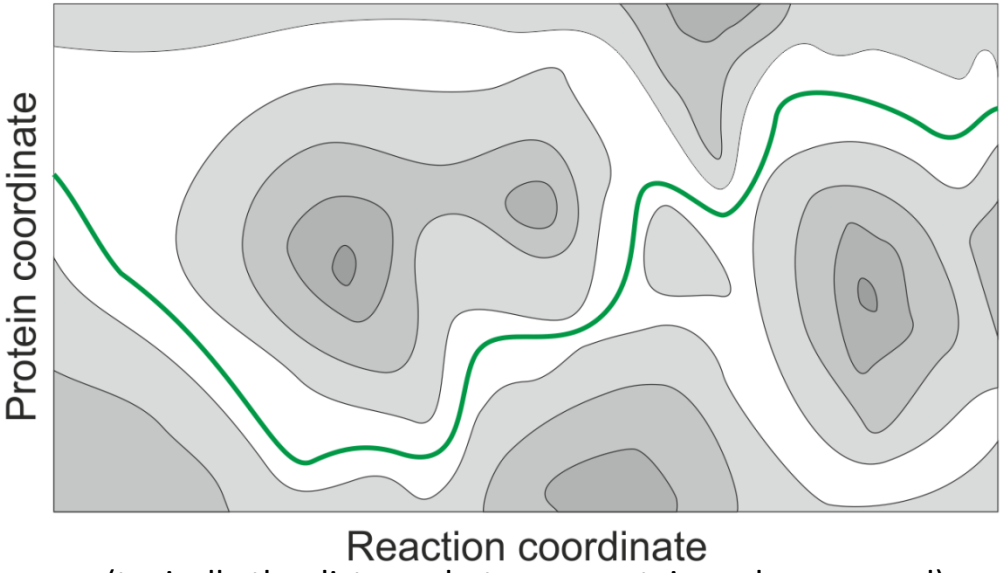
- 4D docking with Autodock vina: **3** protein structures **x** up to **20** docked poses per each **= 60** complexes per compound.
- Iterative convergence in 3D voxel maps similarity and Linear Interactions Energy (LIE) channels
- 4-parameters LIE (Lennard-Jones and Coulomb deltas for compound and binding site)
- Similarity in MD confirmation for the best pose per each compound: 8x replications of 12 ns equilibration + 4 ns production runs.



It is overwhelming to carry out expensive MD study for each compounds in large chemical database, but it is possible to do MD on a subset of compounds and then train AI to quickly predict MD outcome for the entire database.

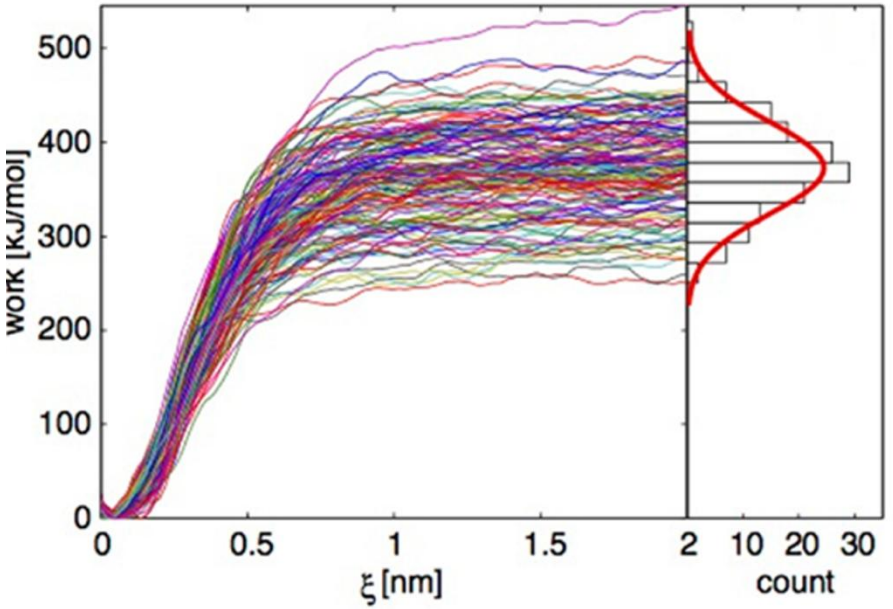
Combination of Jarzynski non-equilibrium pulling and umbrella sampling for free energy estimations.

Umbrella sampling have an excellent “curving” adaptation to relief of energy surface but it is only one run and thus lacks “routes coverage”.



(typically the distance between protein and compound)

Forced pullings are big in number and have good “routes coverage”, but each pulling is short and looks like a straight line over the energy surface.



Zoete et al, 2013

Note. With folded biological molecules it is NOT allowed to use temperature as a factor to pass over the barriers in orthogonal dimensions.

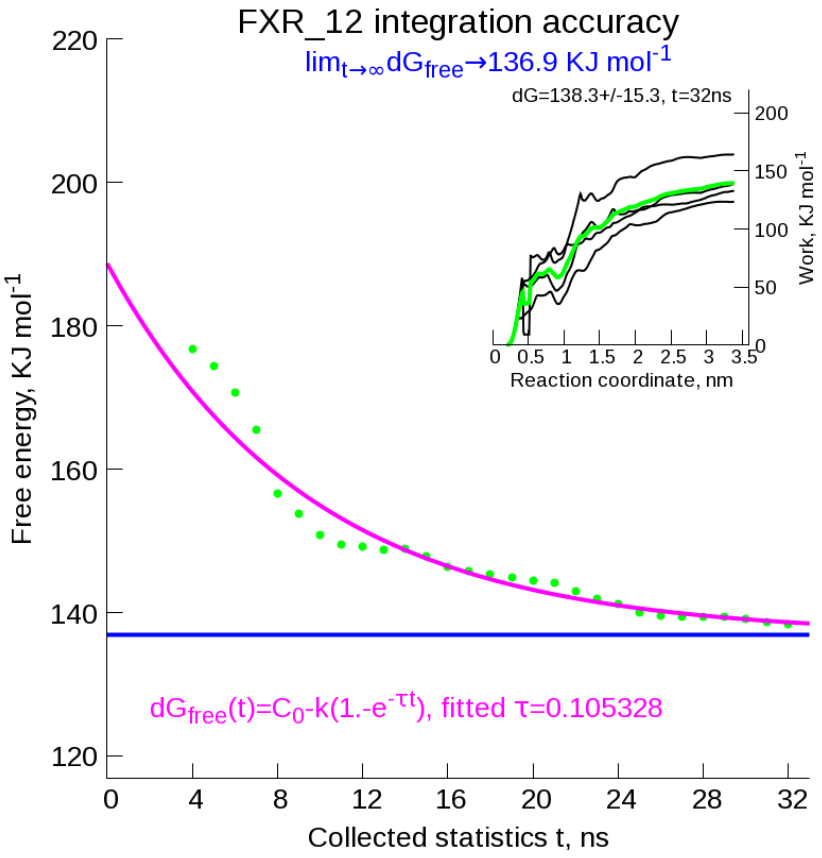
Our improved method combine routes coverage with accurate trajectory sampling.
Practical parameters: amount of routes 4 (16 desired), sampling time 32 ns (~60 desired).

Protocol **vbzci**

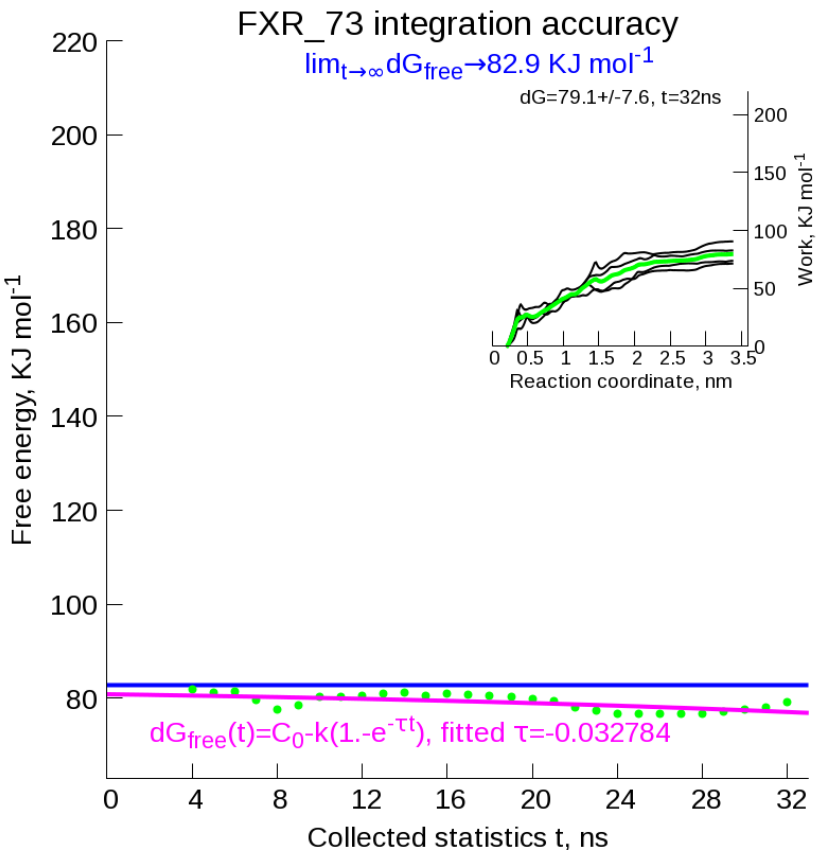
Simple model of time-dependent free energy:

the rate of corrections is proportional to amount of errors.

Slightly overestimated case



Underestimated case (the biggest outlier)



Time-dependent free energy improves standard IFOWON.CO reports as our clients are not only receives cutting edge free energy estimations but also are offered to perform additional computations for the suspected outliers before investing to the chemical synthesis.

Conclusions

The techniques that we found to be valuable at Grand Challenge 2 testing area:

- Involvement of machine learning into drug design
- An optimal combination of Jarzynski pulling and umbrella sampling for free energy estimations

We also identified areas where our technology is now being (further) improved.

In the presentation I am primary focusing on ideas – for more technical details, please, follow up with reading of our planned paper in the special issue of JCAMD.

Many thanks to D3R team for running GC2!!!