Your
partner
of choice
in integrated
research

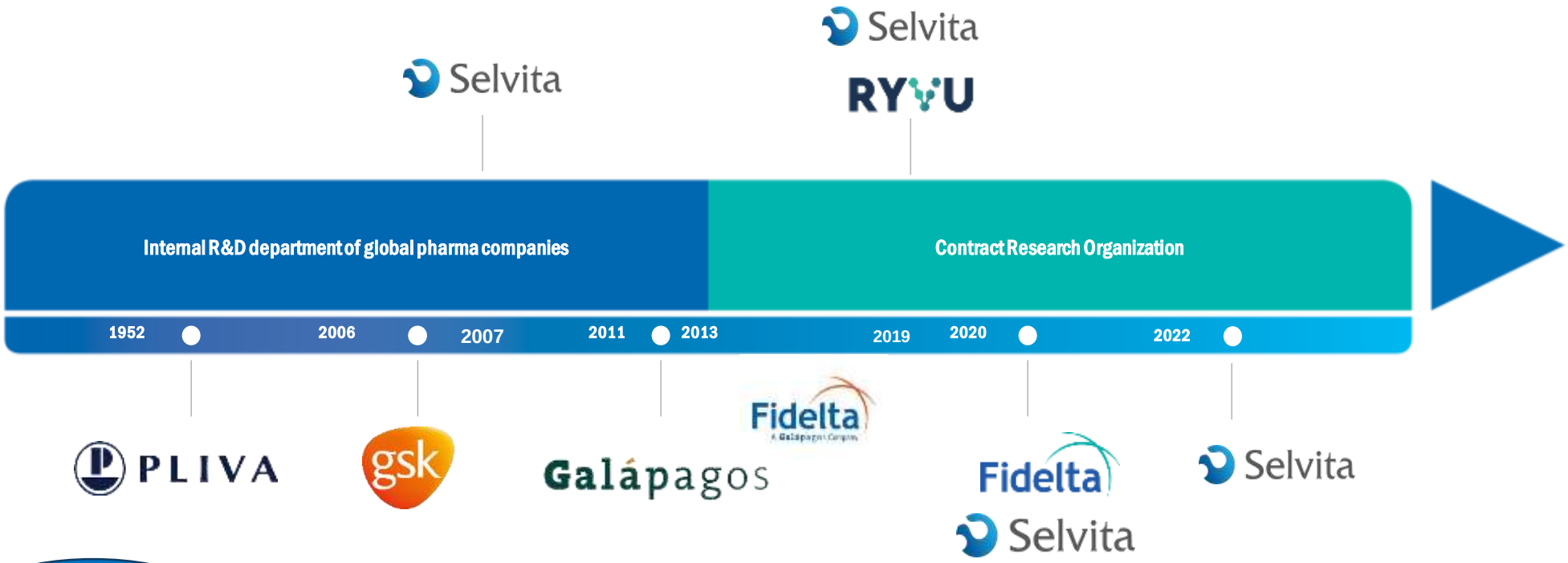
Selvita

HPC impact on preclinical drug discovery

Its importance now and future impacts

Luka Bilić

Contract research organization with longstanding history in drug discovery



60 years of macrolide experience

Combining the best of both worlds

Big pharma and biotech experience with CRO business

Within drug discovery projects there is well designed DMTA cycle

- DMTA stands for Desing, Make, Test, Analyze
- In drug discovery DMTA cycle is an iterative process of:

Designing new compounds

Synthesizing newly designed compounds

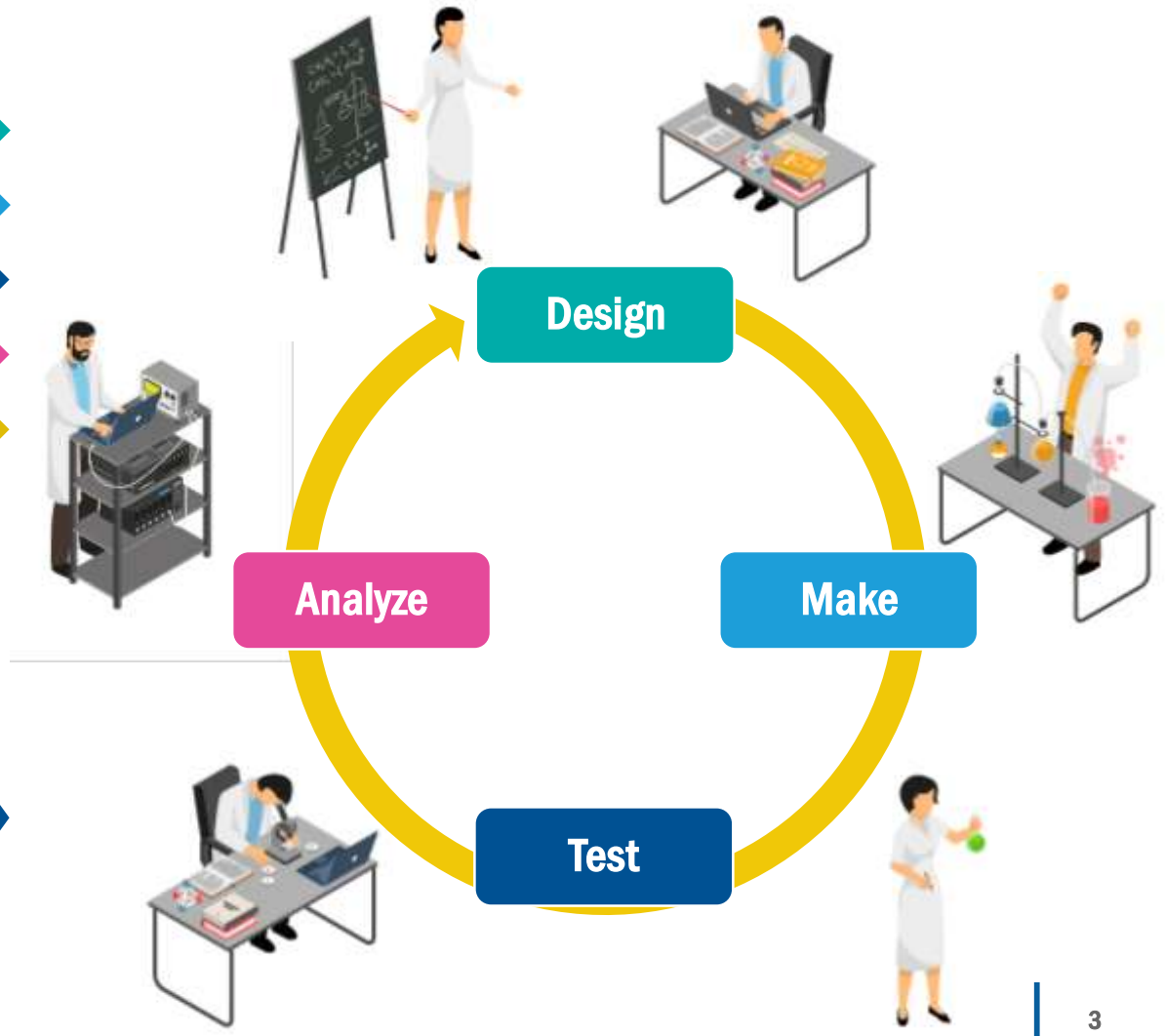
Testing synthesized compounds

Analyzing the results

Incorporating results in the new design cycle

- Depending on the goal, stage of research and individual capabilities DMTA can vary in its details

In the last 2 decades advances in HPC enabled new design strategies and analyses that would not have been possible otherwise



Drug discovery is very expensive business

Selvita

"The mean cost of developing a new drug has been the subject of debate, with recent estimates ranging from $314 million to $2.8 billion" [1]

"After accounting for the costs of high attrition, the median capitalized research and development investment to bring a new drug to the market was estimated at $985.3 million, and the mean investment was estimated at $1335.9 million in the base case analysis." [1]

- Hardware and software development enabled improvements in drug discovery[2]:
  - Shortened drug development by up to 1-year, reduced costs on average by $133 million (30-40% savings for the best-case cost estimate to ~4.5% savings for the worst-case cost estimate)
  - Improved chances of finding hits (molecules that could be developed into drugs)
  - Improved chances of finding new targets (receptors or other cellular/molecular entities which could be targeted to tackle a certain disease)
  - Enabled implementation of novel computational approaches to tackle previously undruggable targets or hard to solve problems

Almost all drug development today employs certain aspects of rational design strategies which necessitate HPC

1. Wouters, O. J., McKee, M., & Luyten, J. (2020). Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA*, *323*(9), 844–853.https://doi.org/10.1001/jama.2020.1166
2. Liu, Tingting et al. "Applying high-performance computing in drug discovery and molecular simulation." National science review vol. 3,1 (2016): 49-63. doi:10.1093/nsr/nww003

HPC in drug discovery

- HPC has many vague definitions; so here is the one I enjoy:

> **"High Performance Computing (HPC) most generally refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve large problems in science, engineering, or business."**

- **What HPC resources do we use and how?**
  - Most of the time we use *In house* HPC resources
  - At Zagreb site there are several dual Xeon Lenovo server racks equipped with Nvidia A100's or Nvidia RTX A4000's with adequate RAM and SSD for our line of work

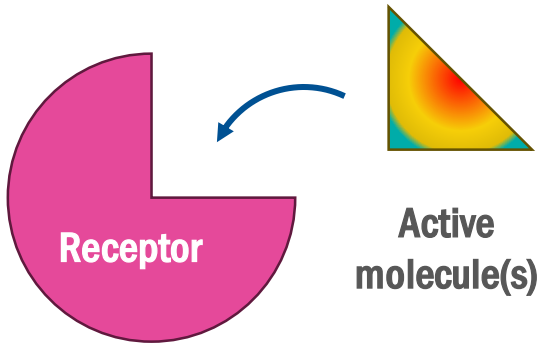
- Larger clients can have their own HPC resources which they can choose to share with us

- Rarely are external HPC resources used due to contracting, privacy and data sharing challenges as well as resource use optimization (financing)
  - Due to project demands we are focused on the cost efficiency - minimum time and resources
  - Fixed number of resources rented on a HPC cluster might be suboptimal scenario due to dynamic nature of DMTA

To understand the HPC hardware use we must understand the general drug discovery pipeline

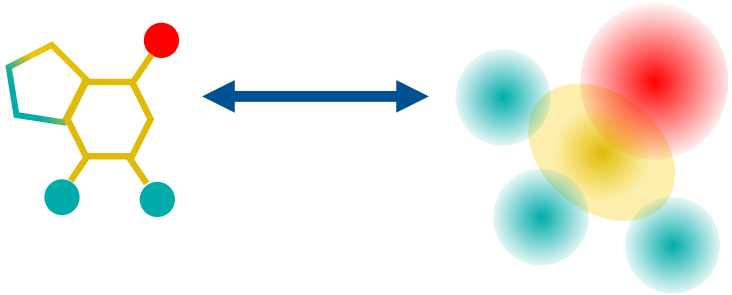
Selvita

Structure-based design
(We know the target receptor structure)



Ligand-based design
(We do not know the target receptor structure, but we know the series of small molecules which show certain activity)

Ligand docking
De novo design
Molecular dynamics
Free energy perturbation (FEP)



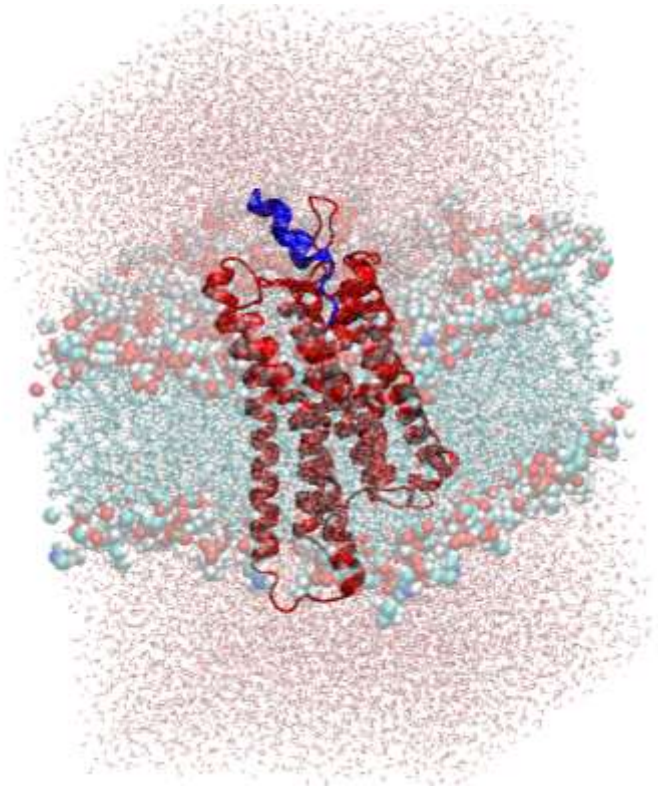
(Qualitative Structure Activity Relationship)
QSAR
Ligand structure information
Pharmacophore modelling
Ligand virtual screening

Lead optimization
(Improving properties that make up an ideal drug candidate: metabolism control, toxicity reduction, solubility optimization etc.)

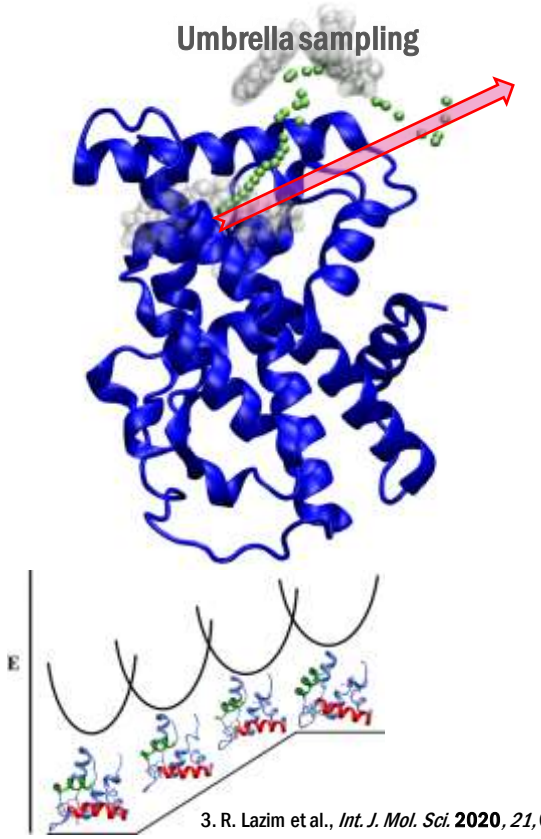
Drug candidate

What in drug discovery merits the use of HPC?

Selvita

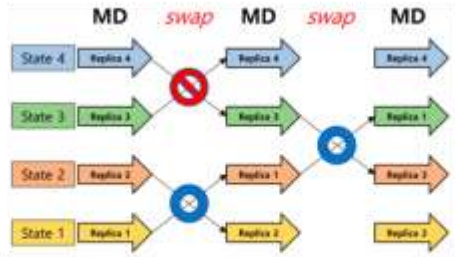


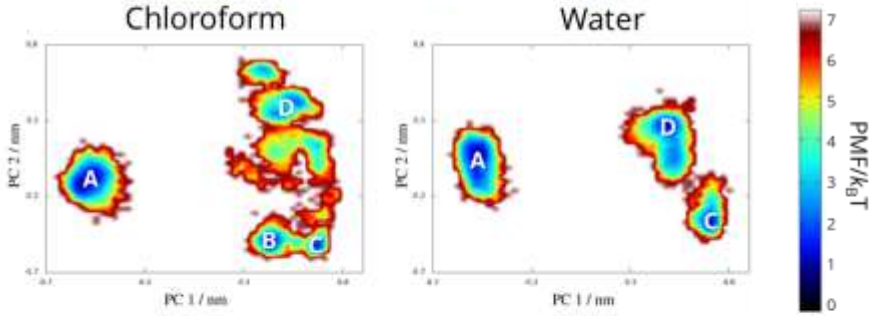
Umbrella sampling

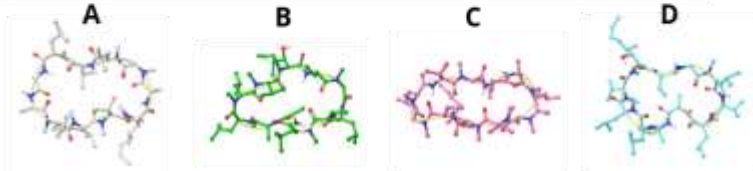


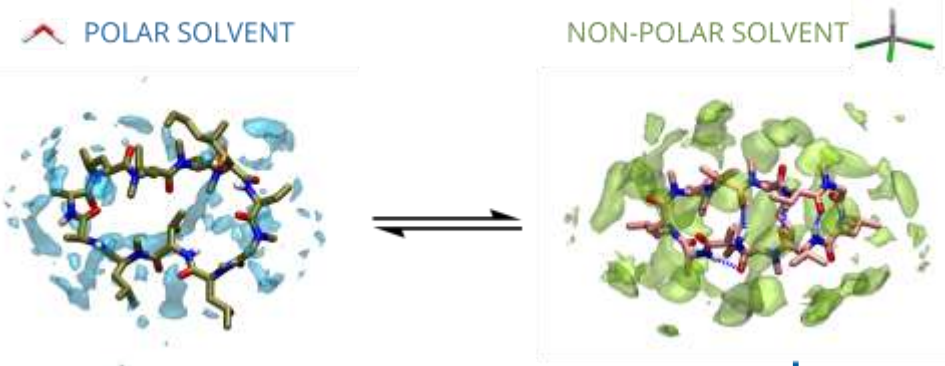
3. R. Lazim et al., *Int. J. Mol. Sci.* **2020**, *21*, 6339.

Replica exchange MD









Large biological targets
- Membrane proteins
- Protein-protein complexes
- Large conformational changes

New chemical modalities:
- Macrocycles, PROTACS, peptides

Complex simulation techniques

- Multiple long simulations
- Large amount of data
- High end GPU for timely execution

HPC hardware usage

Selvita

CPU	GPU	RAM	HDD (SSD)
Often used	Often used	Medium demands	Larger demands
(Usage limited by software licenses)	High end GPU requirements (no current support for tensor cores*)	Larger demands required only for limited type of calculations or simulations (64 GB to 1 TB RAM demands)	Only required for limited type of calculations or simulations (10 or more TB data storage can be exceeded in matter of days)

Fragment or molecular screenings
(docking of 10^6 or more molecules or fragments)

Molecular docking
(Flexible, high precision)

Statistical modelling
(Combination of results obtained from multiple simulations or experiments)

Big data analysis
(When possible)

Molecular dynamics*
(simulating receptor or molecular behavior expected in physiological conditions)

Free energy perturbation*
(Requires many GPU's)

AI/ML in drug development
(when proper molecular data base is available)

***Ab initio* or Quantum mechanical calculations**
(traditionally tends to consume lots of RAM)

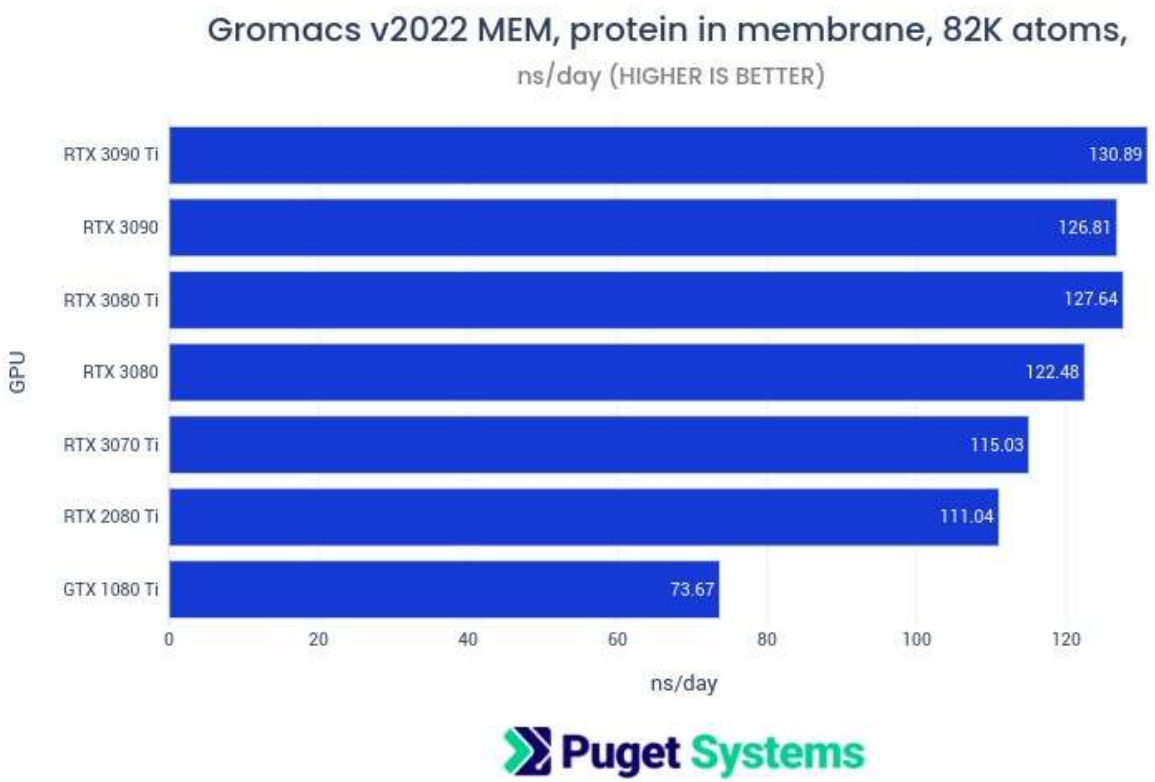
General data analysis
(Especially when dealing with advanced molecular dynamics data analysis)

Molecular dynamics
(can generate large body of useful data in a short time)

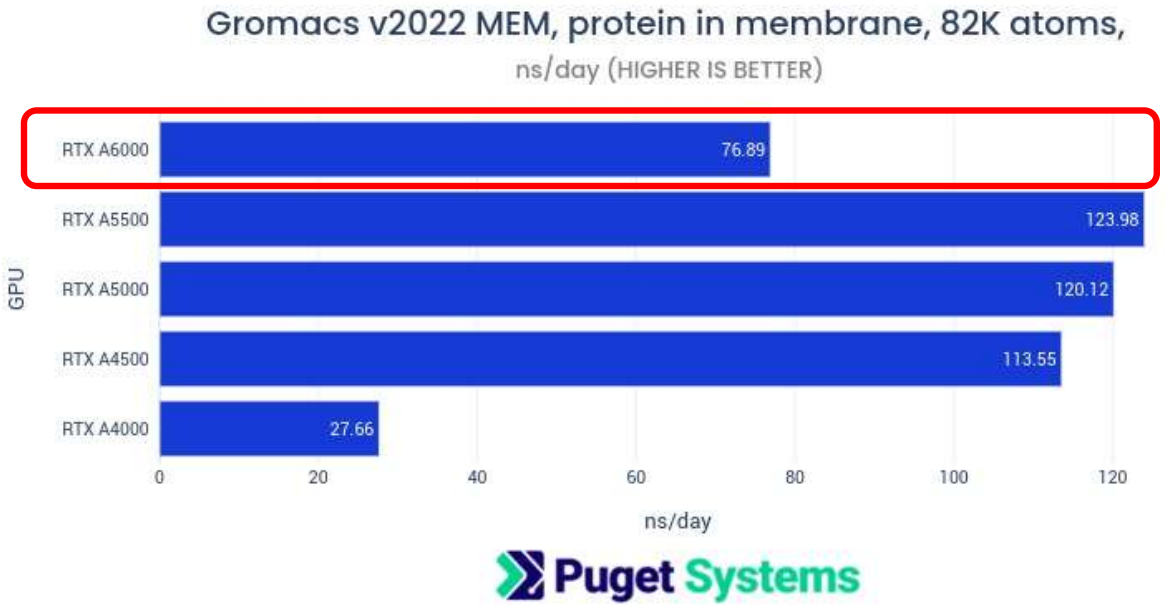
Big data analysis
(When possible)

CUDA support vs Tensor Core support

- This mainly affects molecular dynamics and free energy perturbation simulations
- Higher end GPU does not mean better performance



No support for HPC/workstations



Support for HPC/workstations
But significantly more expensive

4. https://www.pugetsystems.com/labs/hpc/molecular-dynamics-benchmarks-gpu-roundup-gromacs-namd2-namd-3alpha-on-12-gpus-2330/

Future impact of HPC on drug discovery

Selvita

Selvita is expanding:

- New "Selvita global" HPC cluster is on the way with state-of-the-art equipment including Nvidia H100 GPU support for CADD and AI/ML departments
- Individual site locations continue to expand their internal HPC capacities

Molecular dynamics

- Molecular dynamics became readily accessible and continues to open the door in the previously unreachable areas of the drug discovery: **cryptic pockets**, **novel structural data**, **simulations of receptor signaling, protein binding** ***etc.***

Free energy perturbation

- Accurate estimation of **binding affinity** and direct assistance in development of more active compounds
- **Very demanding in terms of GPU power and number** (or time consumption)

AI and Machine Learning

- Automated, faster and more reliable compound optimization in all phases of R&D
- Biomacromolecule structure predictions and structural refinement of possible targets (**Alpha fold** already in action)
- *De-novo* structure design of active compounds with desired properties

THANK YOU FOR ATTENTION!

Luka Bilić

Scientist – Medicinal chemist

CADD group

luka.bilic@selvita.com

/SelvitaGroup

/company/selvita/

SelvitaKrakow