Folding@Home

Actually @ATLAS
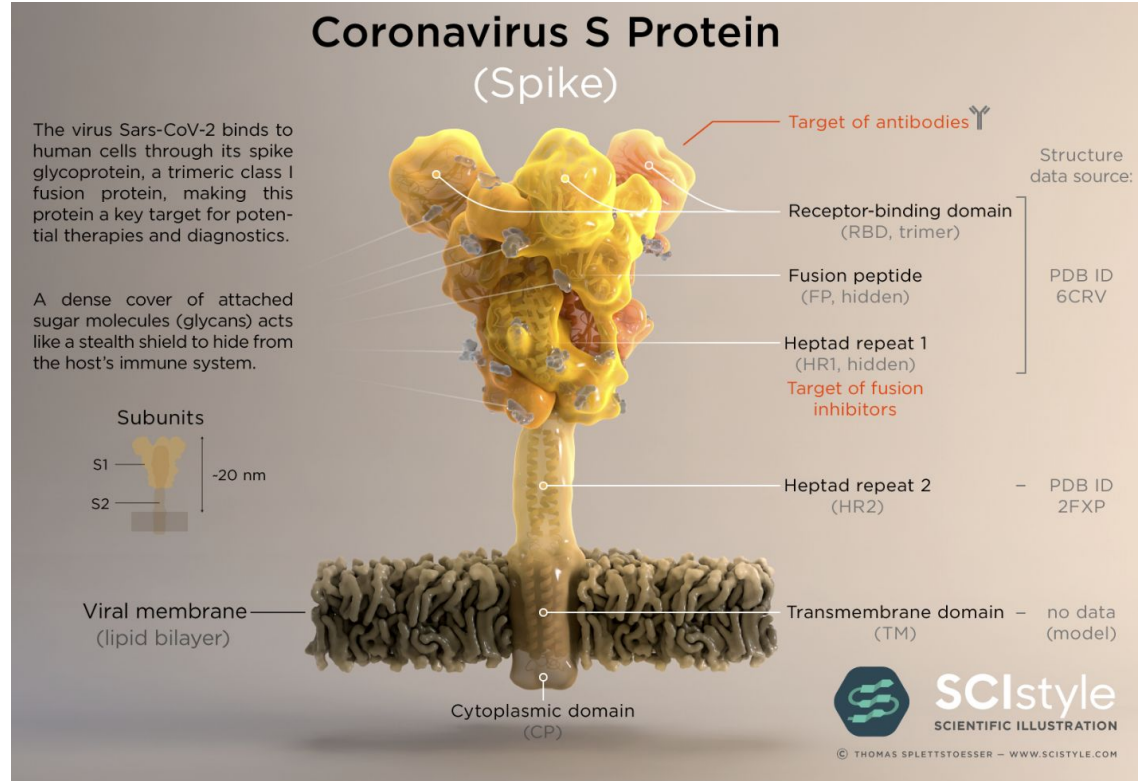
Ilija Vukotic

F@H

A bit of biology:

- RNA & DNA encode proteins (3 bases encode one amino acids).
- Proteins are chains of tens of thousands of amino acids.
- These long chains fold in all kind of ways but in general tend to go to the lowest energy state.
- Protein function is determined by its shape.
- Shape is normally determined by x-ray crystallography. That is an inverse problem so very labor intensive. Any hints are very helpful.
- Calculating the lowest energy conformation is computationally very expensive with million atoms proteins. It gets even harder when a protein works changes during interaction. With a lot of tricks and GPUs still doable.
- [Folding at Home](#) is a very old project that normally runs as a screensaver. So very optimized for Windows and regular desktops, OpenMPI, and OpenCL, checkpoints often, does not need constant WAN connection.

F@H and Covid-19

F@H helps with diseases

- cancer
- neurological
- Infectious

With Covid-19 pandemic, there is a need to identify any potentially "druggable" protein sites on the virus.

Coronavirus S Protein
(Spike)

The virus Sars-CoV-2 binds to human cells through its spike glycoprotein, a trimeric class I fusion protein, making this protein a key target for potential therapies and diagnostics.

A dense cover of attached sugar molecules (glycans) acts like a stealth shield to hide from the host's immune system.

Target of antibodies

Receptor-binding domain
(RBD, trimer)

Fusion peptide
(FP, hidden)

Heptad repeat 1
(HR1, hidden)

Target of fusion inhibitors

Structure data source:

PDB ID
6CRV

Subunits

S1

S2

-20 nm

Heptad repeat 2
(HR2)

– PDB ID
2FXP

Viral membrane
(lipid bilayer)

Transmembrane domain
(TM)

– no data
(model)

Cytoplasmic domain
(CP)

SCIstyle
SCIENTIFIC ILLUSTRATION
© THOMAS SPLETTSTOESSER – WWW.SCISTYLE.COM

F@H scheduling

Calculations are organized in Projects - just like our tasks.

Projects are split in work units (WUs) - equivalent of our jobs.

Different projects have different priorities.

For a finished WU, donor gets Credits.

WUs are of different durations.

They try to optimize Projects Time-To-Complete, just as we are trying to do.

As we have rebrokering, they have it too. Each WU has a timeout and a deadline.

If the timeout is reached, WU gets sent to another donor. If a deadline passes or the other donor finishes the WU, original donor gets a new WU, but gets no credits for the calculation done.

Us and F@H

CERN asked WHO how to best contribute fight against pandemic, and they recommended F@H.

Ivan G. submits tasks that run F@H at different PQs. Each job is one WU. We donate both CPU and GPU resources.

Accounting is under [CERN team](#).

CPU resources are listed as [ATLAS_CPU](#).

Individual CPU contributions can be seen [here](#), or in ES, BigPanda (gShare=COVID).

GPU resources are listed according to PQ ([ANALY_MWT2_GPU](#), [ANALY_MANC_GPU_TEST](#))

US sites also contribute via [OSG](#).

UChicago contributes with opportunistically run F@H on the [ATLAS ML platform](#).

If you have idling GPUs here my [dockerhub repo](#), my [k8s deployments](#), it can be started with one line. The only requirement is that a node can access port 80 or 8080.

Statistics

Donor	Credit	WU
ATLAS_CPU	113M	180k
UC ML Platform	30M	4.5k
BNL_HPC_CPU	11M	2.3k
ANALY_MWT2_GPU	4M	372
OSG	56M	118k

Credit/WU is obviously very different.

Two options:
- Our CPUs get very low priority jobs
- We don't do them quickly enough so don't get awarded credits.

Proposal - reduce our CPU contributions and employ GPUs as much as we can. (SLAC, BNL,...)