In silico studies of the mutant protein causing cystic fibrosis

Tamás Hegedűs

tamas@hegelab.org

MTA-SE Molecular Biophysics Reasearch Group
Dept. Biophysics and Radiation Biology, Semmelweis University

Acknowledgement

www.hegelab.org

Tordai Hedvig
László Laura
Csizmadia Georgina

Tóth Ágota
Spagina Zoltán

Kellermayer Miklós

SE Biofizikai és Sugárbiológiai Intézet
MTA-SE Molekuláris Biofizikai Kutatócsoport

Sarkadi Balázs

MTA TTK, Enzimológiai Intézet
MTA-SE Molekuláris Biofizikai Kutatócsoport

Gergely L. Lukacs, Naoto Soya, Ariel Roldan

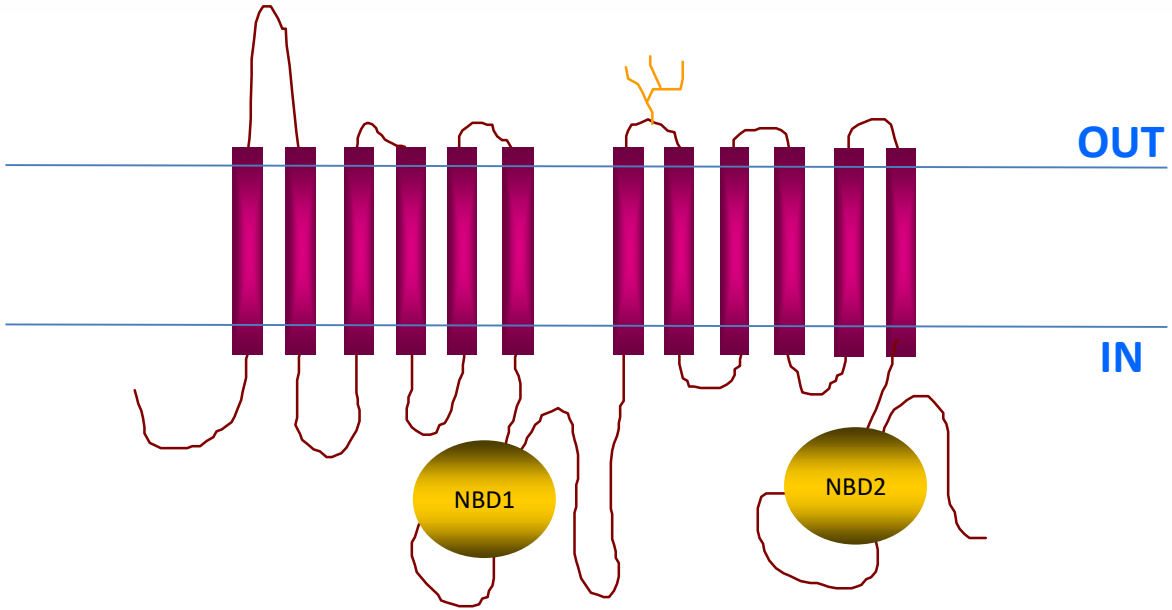
Dept. Physiology, McGill University

FUNDING:
MTA Bolyai Fellowship
NKIFH K 111678

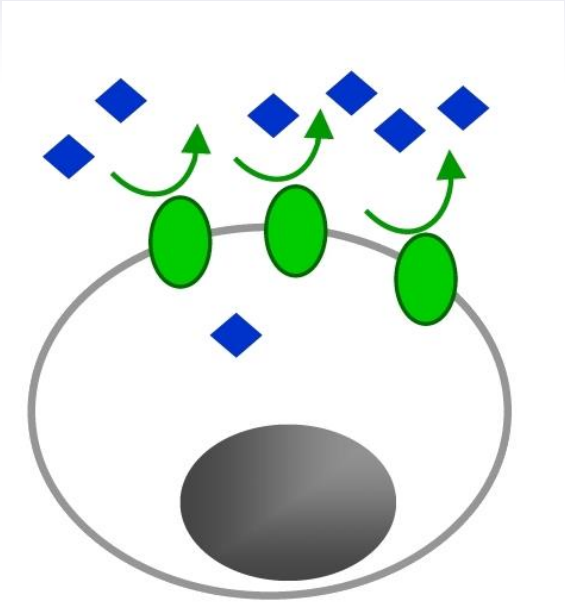
- NIIF HPC
- Wigner GPU laboratórium

  Barnaföldi Gergely
- Bartha Ferenc (Szeged)

ABC (ATP Binding Cassette) transmembrane proteins



Transport of xenobiotics and toxic molecules



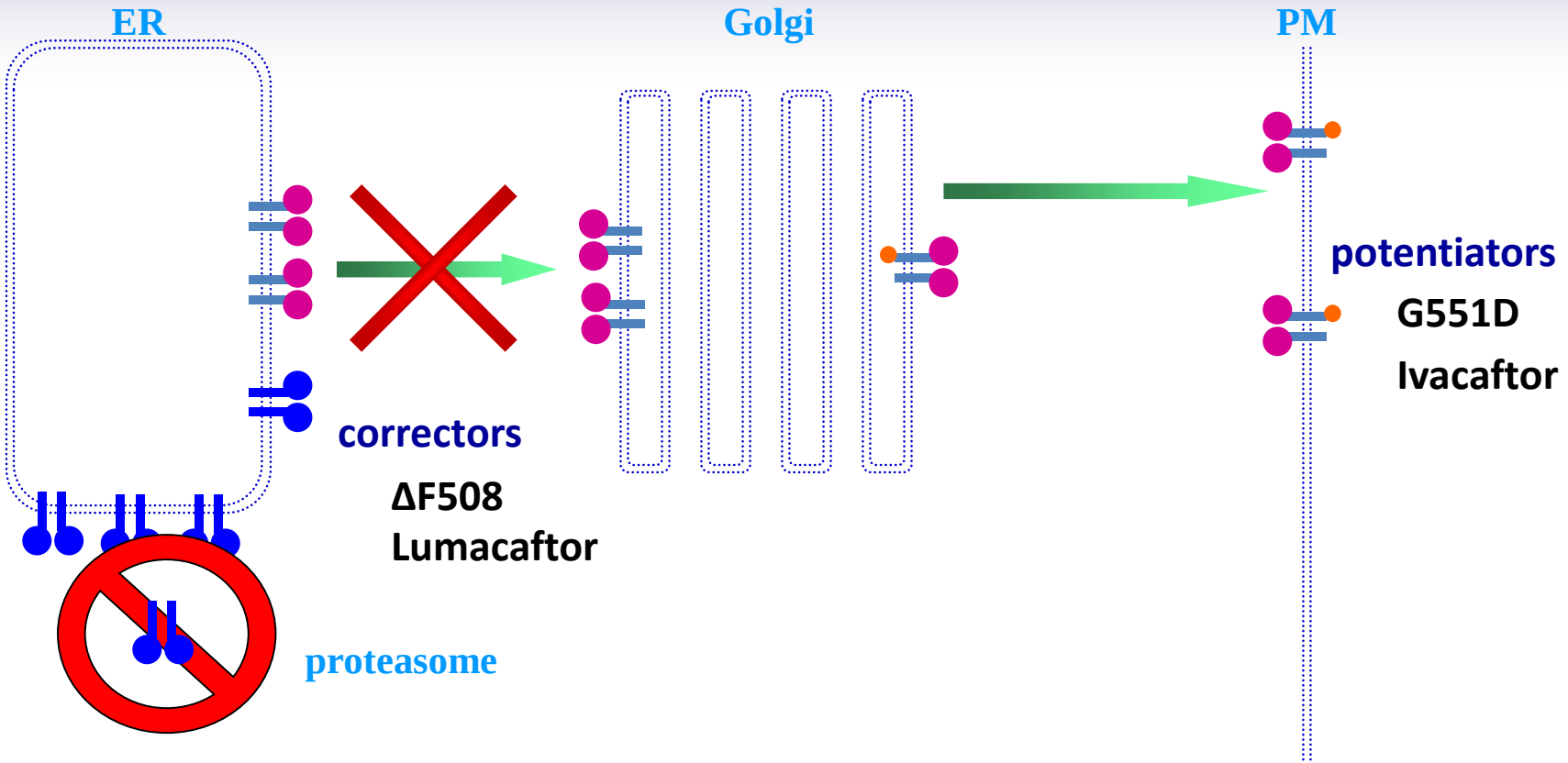
Keywords:

- cancer multidrug resistance
- drug pharmacokinetics
- drug-drug interactions, etc.

Cystic Fibrosis (CF)

- **monogenic disease** (~1:2,500)
- **high mortality and morbidity**
- **affected water and salt homeostasis**
- **cloning the *cftr* gene:** Riordan *et al.* 1989 *Science* **245**:1066-73
- **over 2,000 mutations are known**
- **ΔF508 is present in over 80% of patients**

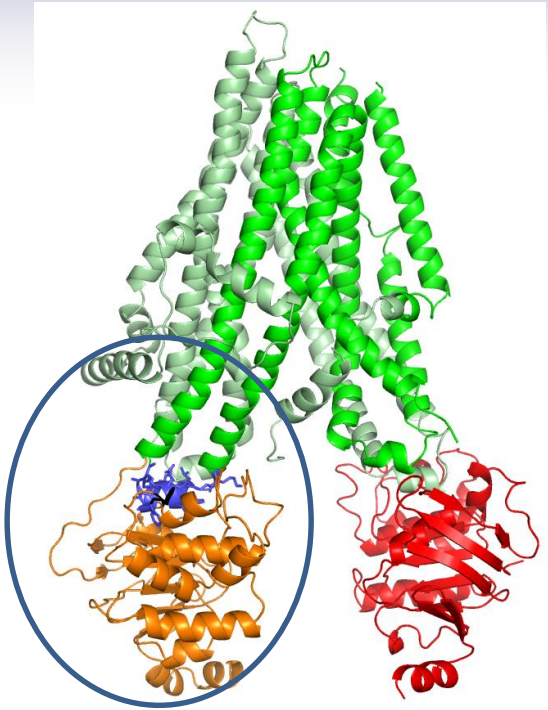
Maturation of the CFTR protein

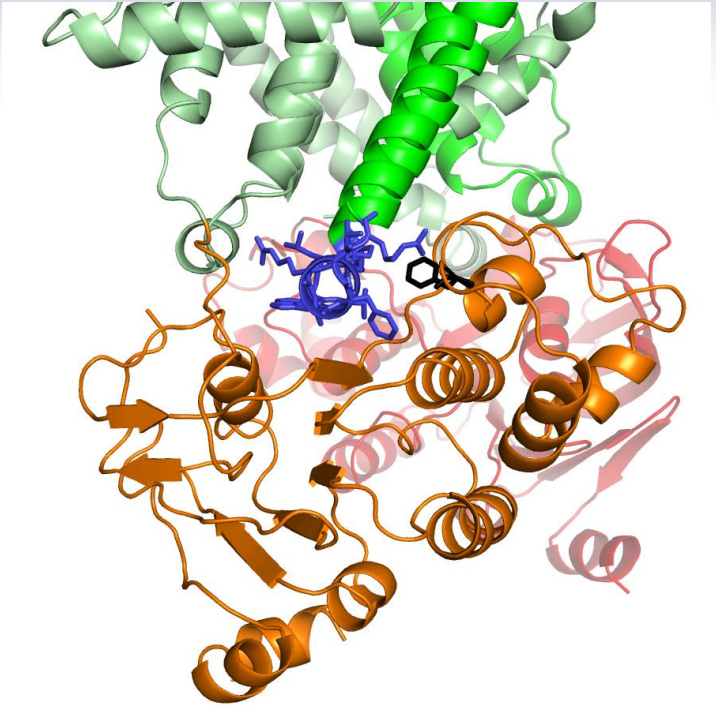


Structural background of ΔF508

domain-domain interactions

hCFTR (PDBID:5U71)





Du *et al.* Nat Struct Mol Biol. **2005** Jan;12(1):17-25.

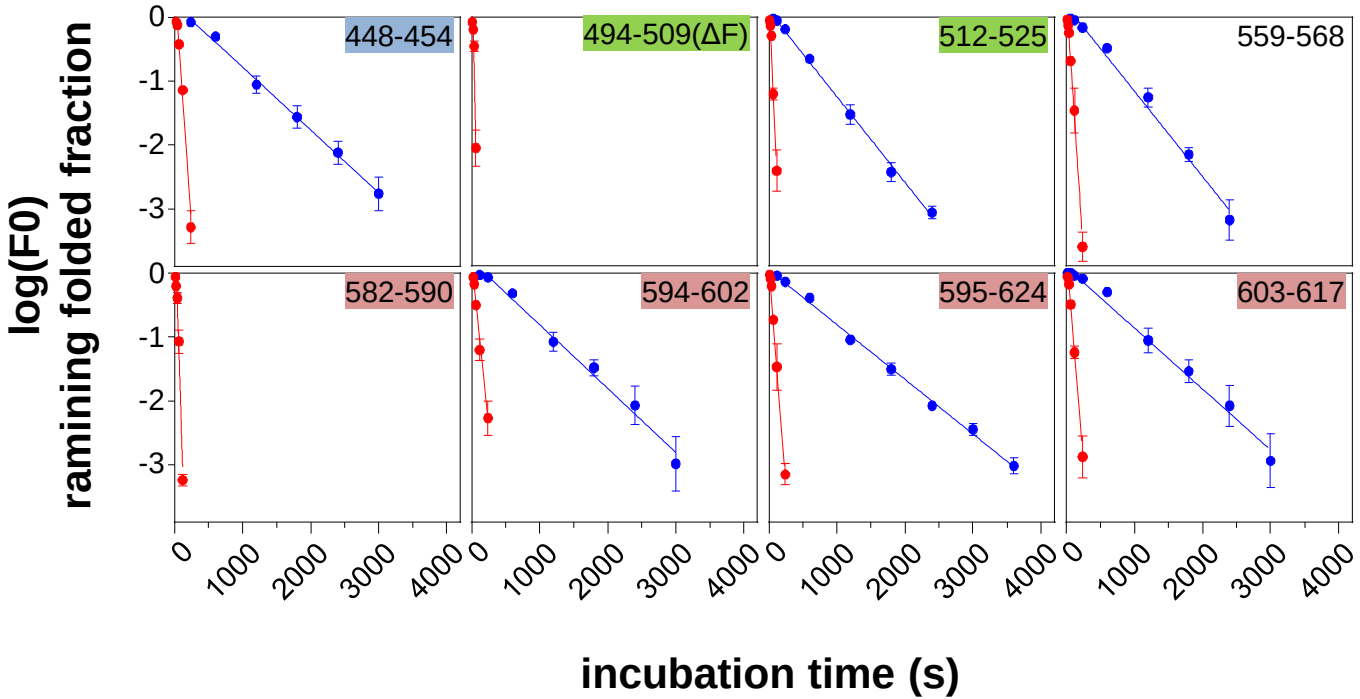
Serohijos *et al.* Proc Natl Acad Sci U S A. **2008** Mar 4;105(9):3256-61.

Structural background of ΔF508

folding and stability of NBD1

NBD1

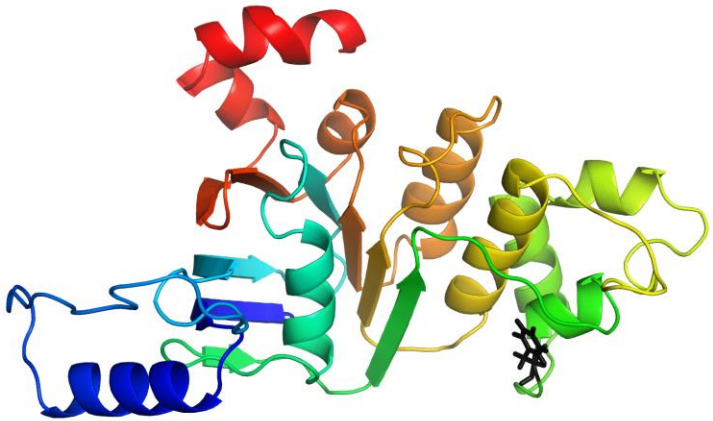
Experiment: Hydrogen-Deuterium Exchange (HDX)



In silico study of NBD1 destabilization

1. Molecular dynamics (MD) simulations
   - WT and mutants
   - 100-100 ns
   - *all-atom force field*

2. Analysis of motions in NBD1
   - correlations in motions (pairwise)
   - building a graph
   - analysis of the graph



Correlation in motions

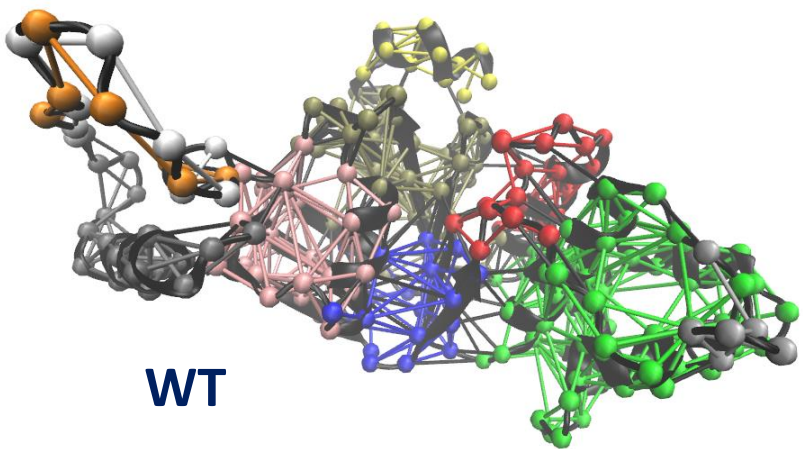
Nodes: amino acids

Edges:

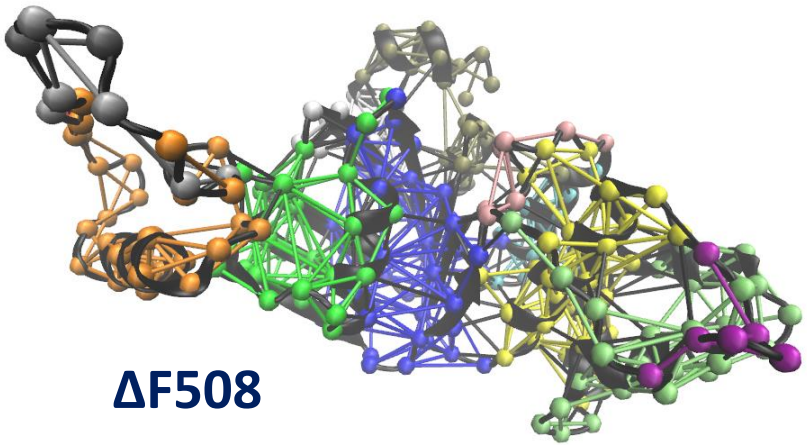
- **displacement Vector Correlation Coefficient**
- **Contact over 75% throughout the simulation**

$$\mathrm{VCC} = \frac{\langle (\mathbf{A} - \langle \mathbf{A} \rangle)(\mathbf{B} - \langle \mathbf{B} \rangle) \rangle}{\sqrt{\langle (\mathbf{A} - \langle \mathbf{A} \rangle)^2 \rangle \langle (\mathbf{B} - \langle \mathbf{B} \rangle)^2 \rangle}}$$

Community analysis (e.g. critical nodes)



WT



ΔF508

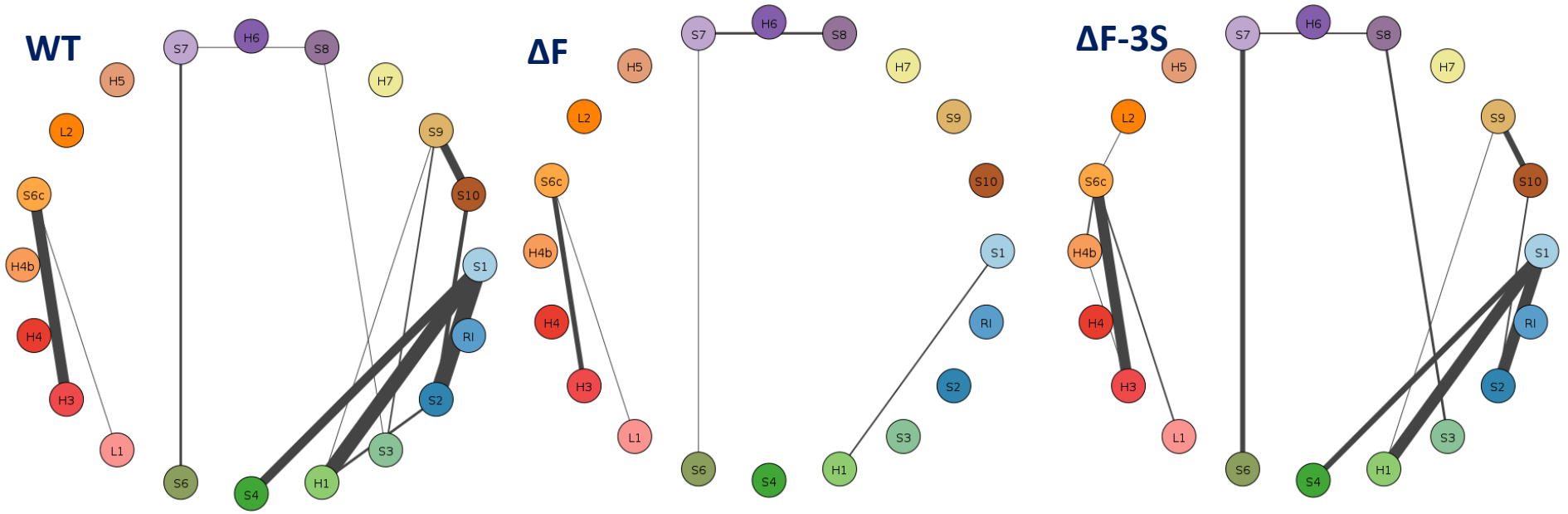
Distance Correlation Coefficient

G. J. Szekely *et al.* (2007), Annals of Statistics, 35 (6): 2769–2794.

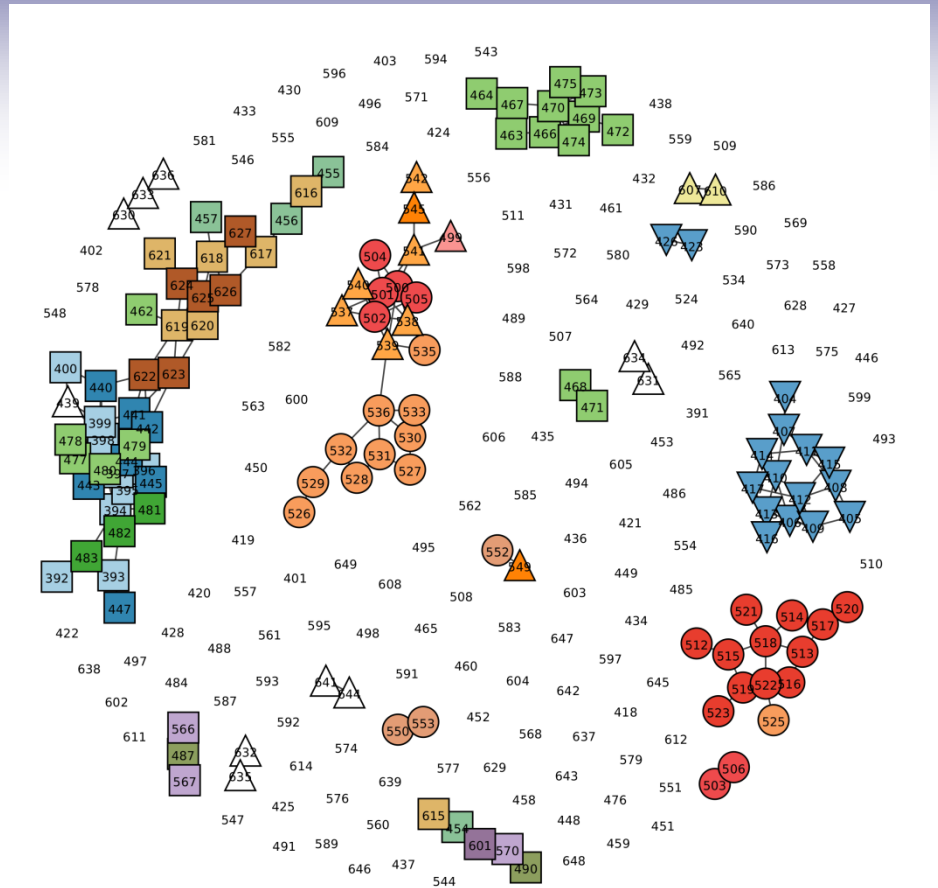
$$\mathrm{DiCC} = \frac{\nu(\mathbf{A}, \mathbf{B})}{\sqrt{\nu(\mathbf{A}, \mathbf{A})\nu(\mathbf{B}, \mathbf{B})}}$$

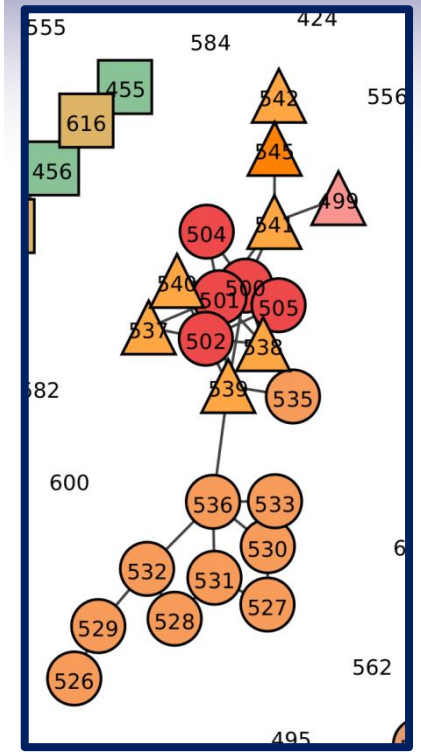
Nodes: secondary structural elements

Edges: if DiCC > 0.85



Identification of critical amino acids

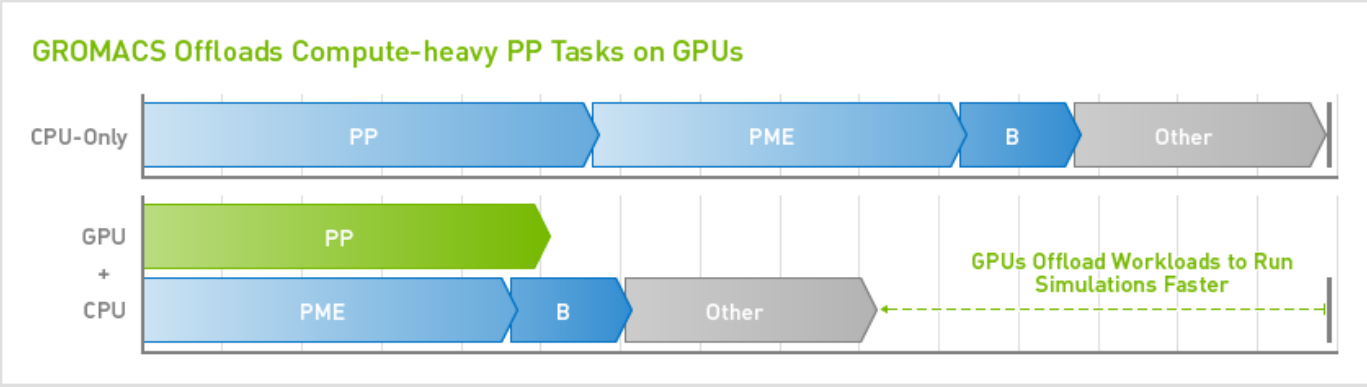




Utilization of GPUs #1

Running MD simulations, GROMACS

http://www.nvidia.com/object/running-jobs-in-gromacs.html



PP: Calculate short range non-bonded forces or particle-particle (PP) interactions

PME: Calculate an approximation for the long range part of long range non-bonded forces

- intensive internode communication
- 3DFFT

Moving PME to GPU worth only if job runs on a single node

Utilization of GPUs #2

Analyzing the trajectories, DiCC

$$\mathrm{DiCC} = \frac{\nu(\mathbf{A}, \mathbf{B})}{\sqrt{\nu(\mathbf{A}, \mathbf{A})\nu(\mathbf{B}, \mathbf{B})}}$$

$$\nu(\mathbf{A}, \mathbf{B}) = \sqrt{\frac{1}{n^2} \sum^{ij} \alpha_{ij}\beta_{ij}}$$

where

$$\alpha_{ij} = a_{ij} - a_{i.} - a_{.j} + a_{..} \qquad (2)$$

The following steps are needed to calculate α_{ij} from $\{\mathbf{A}\}$.

1. Build the $n \times n$ matrix, a, from $\{\mathbf{A}\}$, where a_{ij} is the distance between the ith and jth entries of $\{\mathbf{A}\}$: $a_{ij} = |\mathbf{A}^i - \mathbf{A}^j|$
2. Average the rows of $\mathbf{a}$: $a_{i.} = (1/n)\sum_j a_{ij}$
3. Average the columns of $\mathbf{a}$: $a_{.j} = (1/n)\sum_i a_{ij}$
4. Average all elements of $\mathbf{a}$: $a_{..} = (1/n^2)\sum_{ij} a_{ij}$
5. Build the $n \times n$ matrix $\boldsymbol{\alpha}$ from $\mathbf{a}$ where $\alpha_{ij} = a_{ij} - a_{i.} - a_{.j} + a_{..}$

My best case scenario:

250 amino acids
(250x250 symmetric matrix)
n(frames) = len(vector) = 8000

approx 120Gb RAM
approx 1 hour run on

- **Intel 8 cores**
- **AMD 32 cores**

Summary

- We employ *in silico* tools to dissect important changes in protein dynamics and identify critical residues for drug development.
- Accelerating running and analysis of simulations are critical.
- The most challenging steps are the interpretation and presentation of the results.

Thanks for your attention!

Cystic fibrosis