Accelerating molecular docking on Graphics Processors

Simon McIntosh-Smith

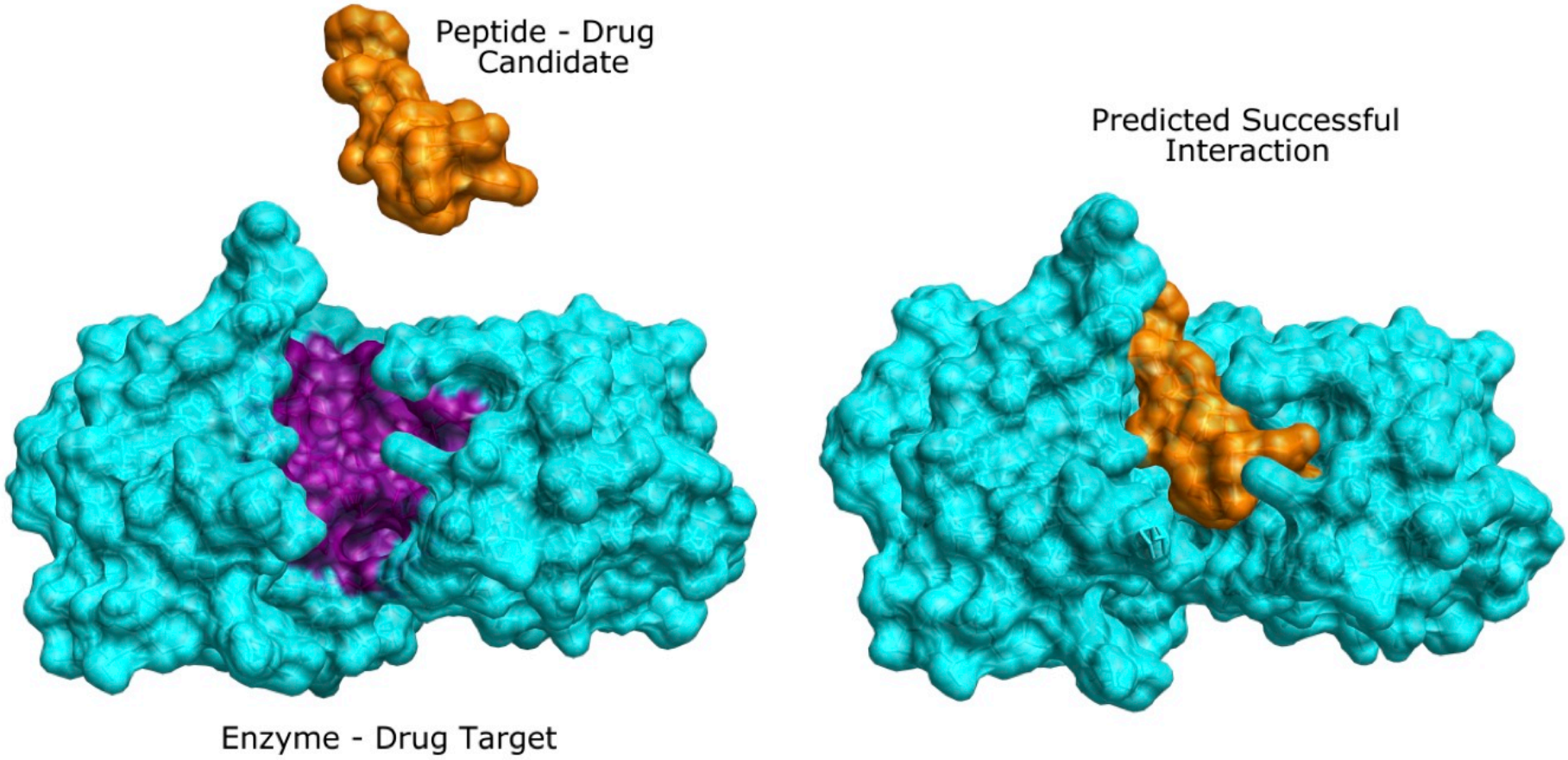
Richard B. Sessions

University of Bristol, UK

simonm@cs.bris.ac.uk, R.Sessions@bristol.ac.uk

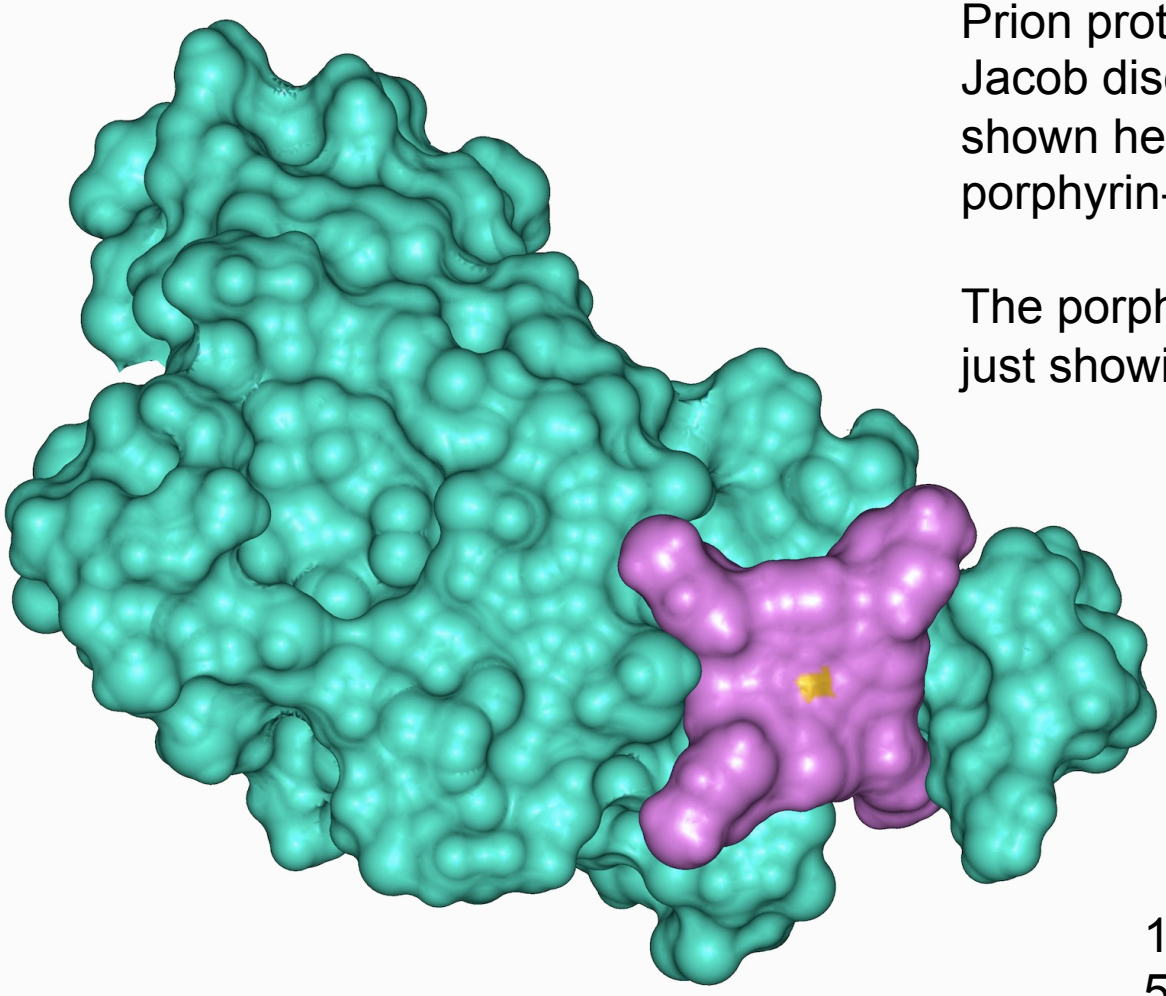
University of BRISTOL

Drug docking examples: Elastase inhibitors



University of BRISTOL

Prion disease



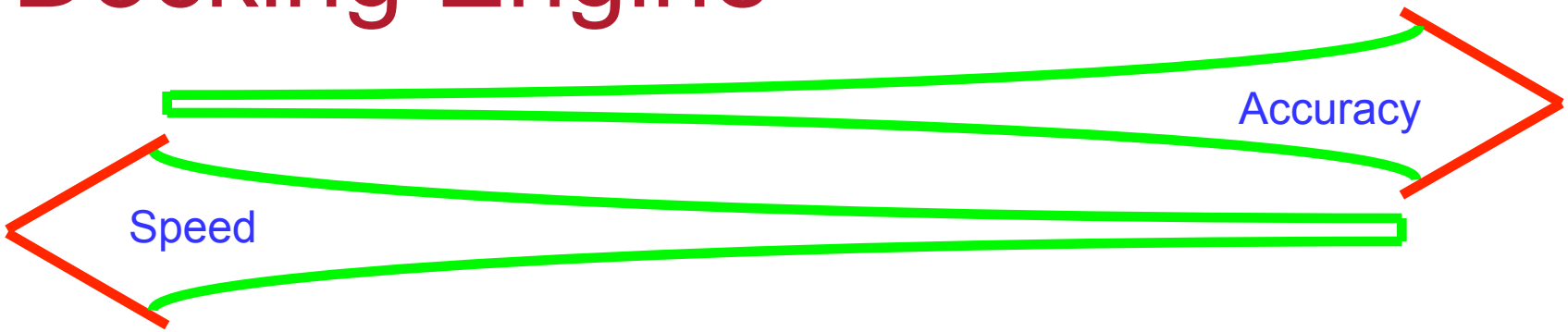
Prion protein behind Creutzfeld-Jacob disease in humans and shown here binding with a (pink) porphyrin-based ligand

The porphyrin's bound iron ion is just showing in yellow

1,719 atoms in the protein
53 atoms in the ligand

University of BRISTOL

BUDE: Bristol University Docking Engine



Typical docking scoring functions	Empirical Free Energy Forcefield **BUDE**	Free Energy calculations MM[1,2] QM/MM[3]
Entropy:		
solvation	**No**	**Yes**	**Yes**
configurational	**Approx**	**Approx**	**Yes**
Electrostatics	**?**	**Approx**	**Yes**
All atom	**No**	**Yes**	**Yes**
Explicit solvent	**No**	**No**	**Yes**

University of BRISTOL

1. MD Tyka, AR Clarke, RB Sessions, J. Phys. Chem. B 110 17212-20 (2006)
2. MD Tyka, RB Sessions, AR Clarke, J. Phys. Chem. B 111 9571-80 (2007)
3. CJ Woods, FR Manby, AJ Mulholland, J. Chem. Phys. 128 014109 (2008)

Empirical Free Energy Function (atom-atom)

$$\Delta G_{\text{ligand binding}} = \sum_{i=1}^{N_{\text{protein}}} \sum_{j=1}^{N_{\text{ligand}}} f(x_i, x_j)$$

Parameterised using experimental data†

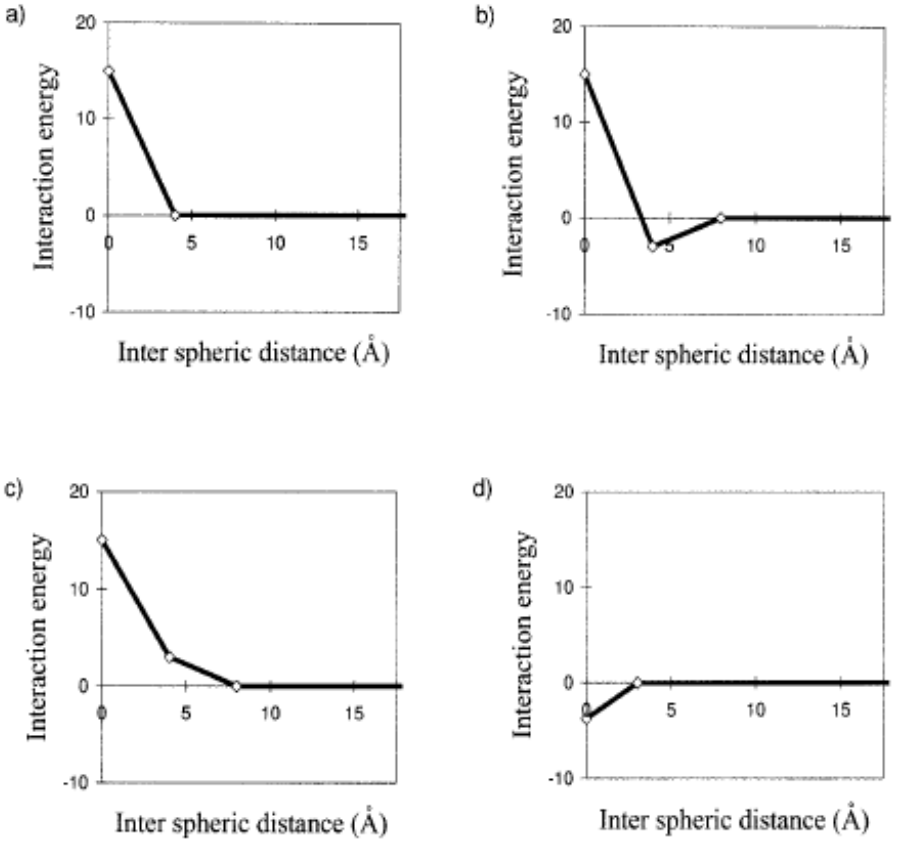
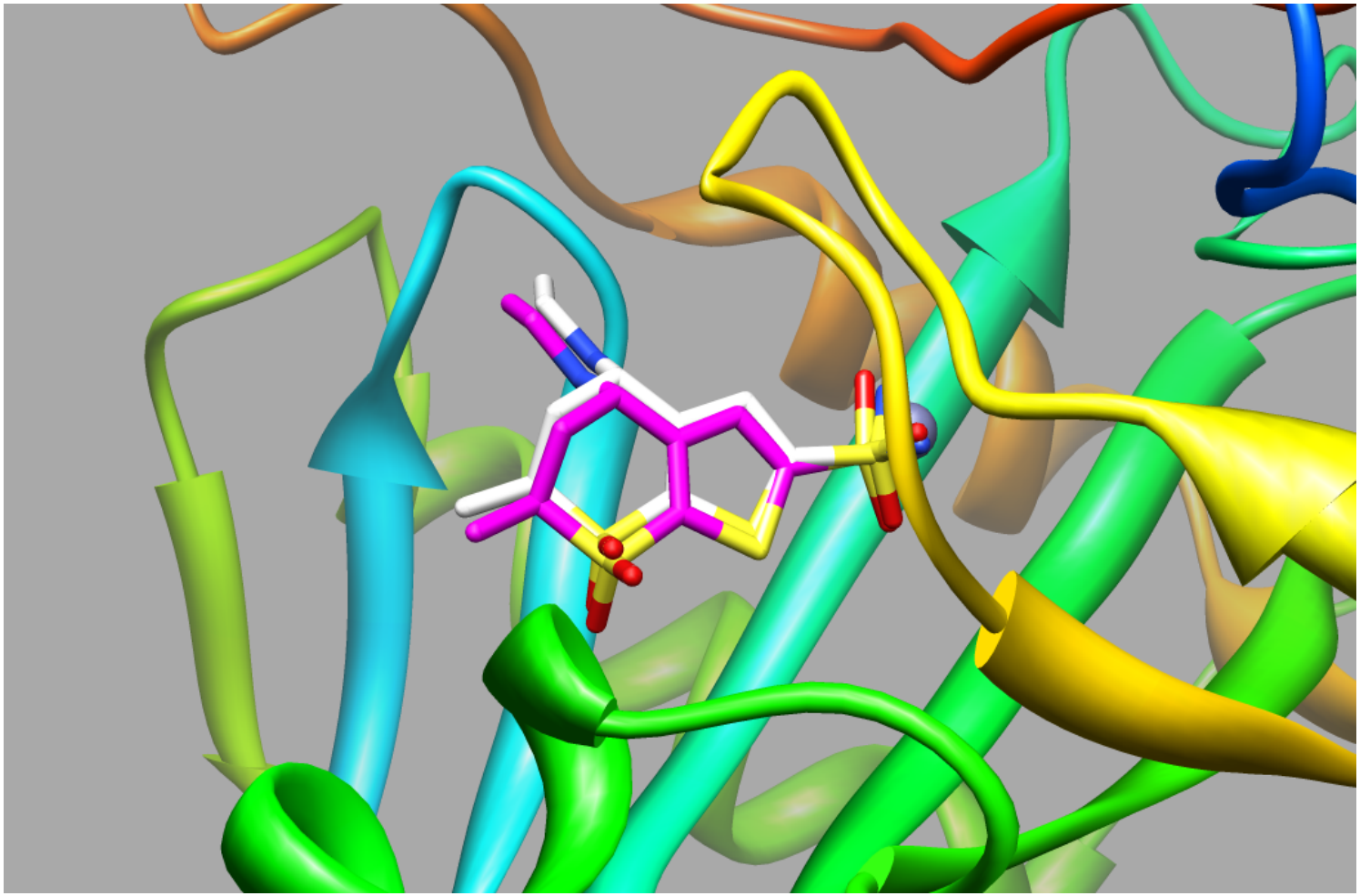


Fig. 1. Inter-residue sphere-sphere interaction energy functions of the force field. a: Between two polar spheres, or between a backbone sphere and any other non hydrogen-bonding sphere. b: Between two non-polar spheres. c: Between a non-polar sphere and a polar sphere. d: Between a hydrogen bond donor sphere and a hydrogen bond acceptor sphere.

University of BRISTOL

† N. Gibbs, A.R. Clarke & R.B. Sessions, "Ab-initio Protein Folding using Physicochemical Potentials and a Simplified Off-Lattice Model", Proteins 43:186-202,2001

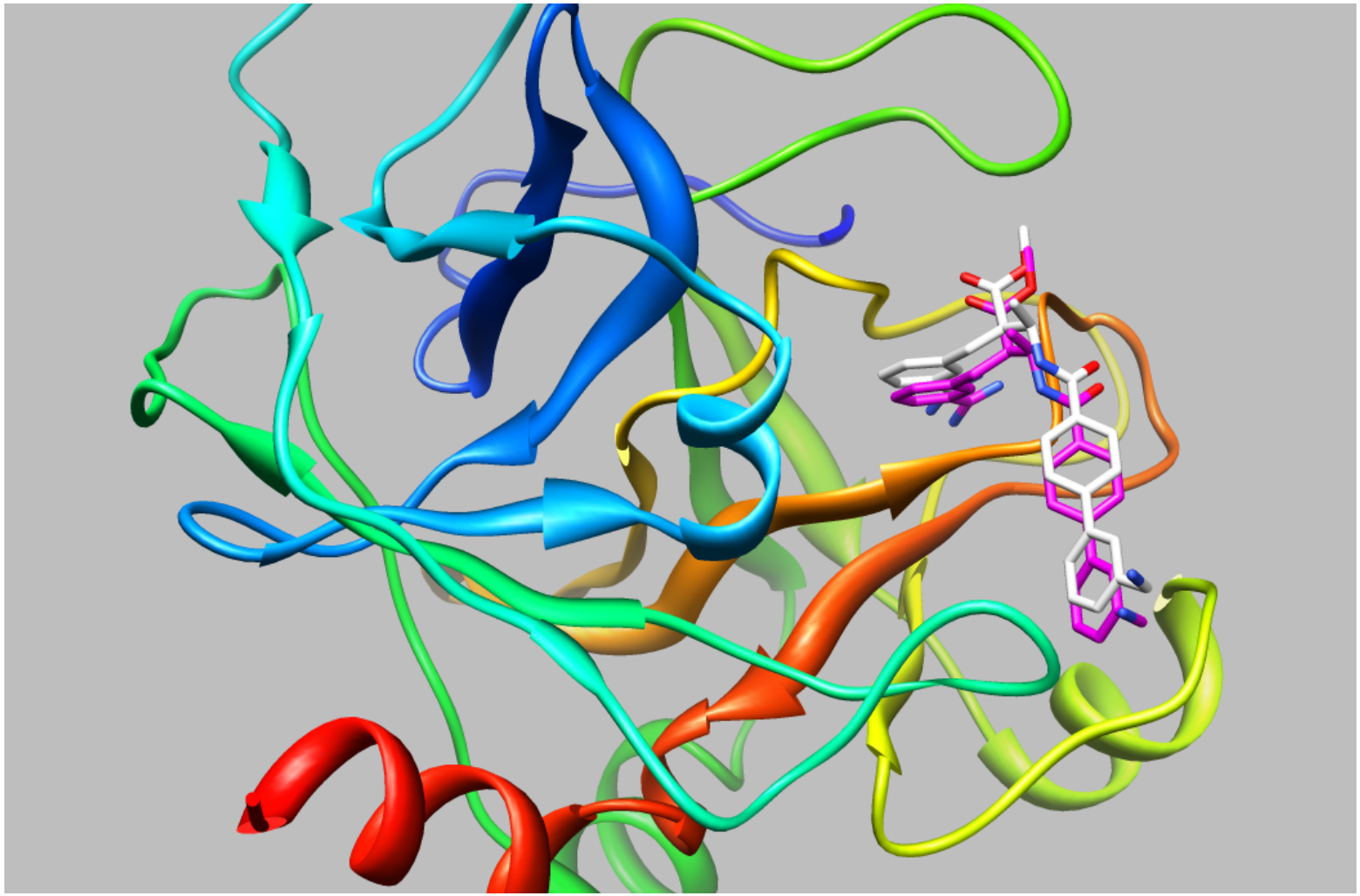
Redocking into Xray Structure



University of BRISTOL

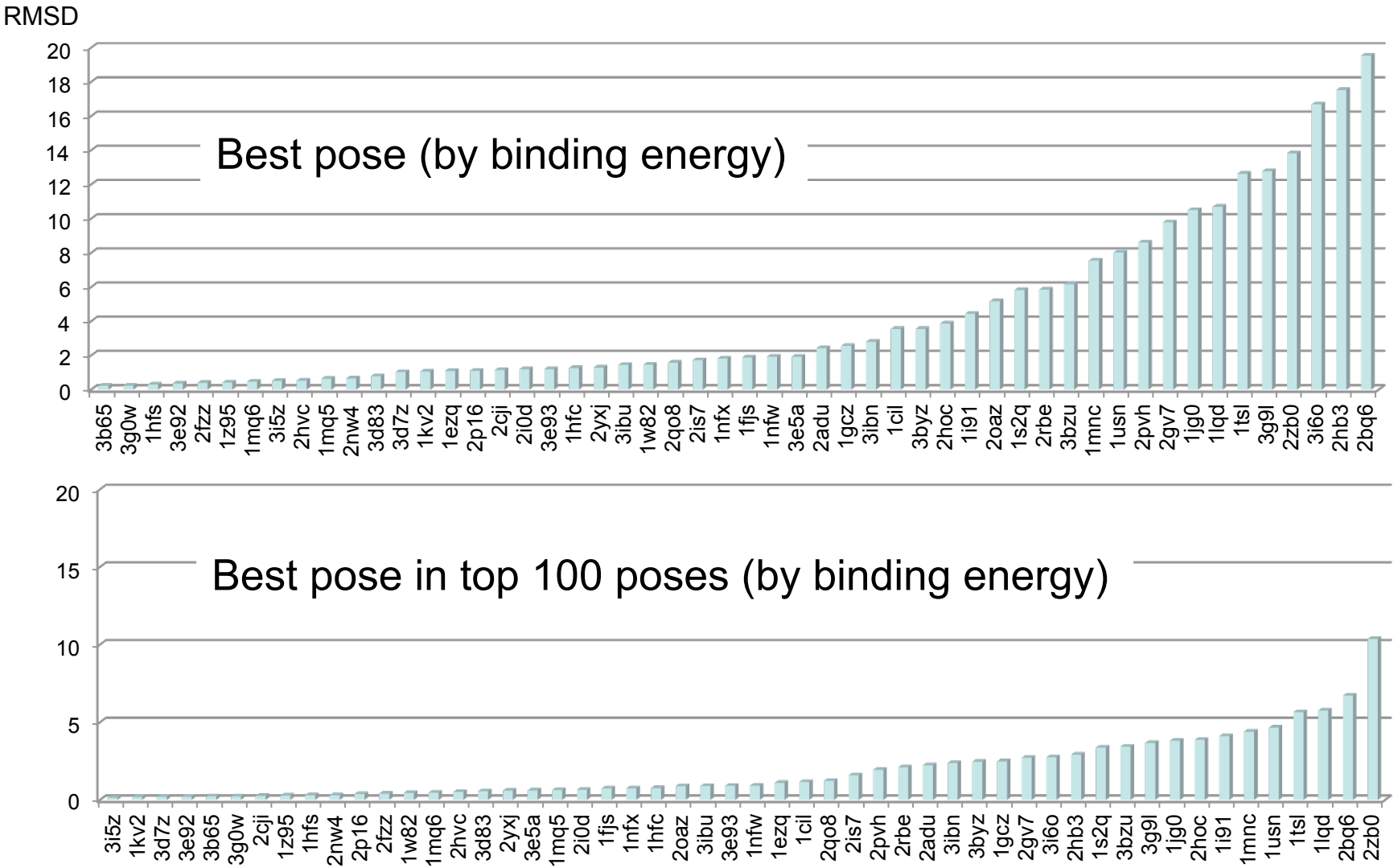
1CIL (Human carbonic anhydrase II)

Another example



1EZQ (Human Factor XA)

Self docking results (Ligand Binding Database)

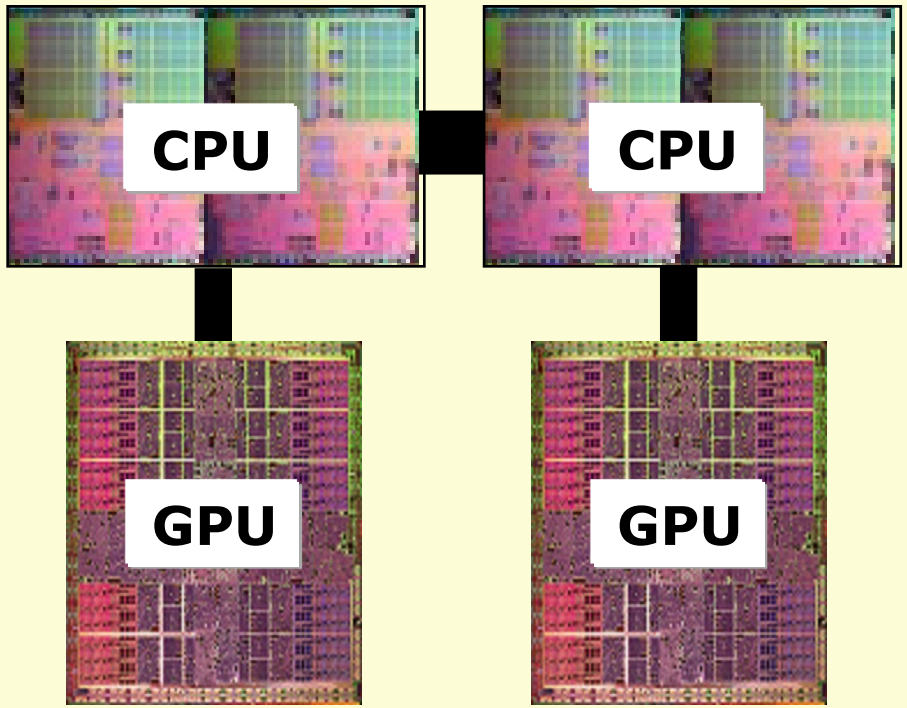


University of BRISTOL

What are GPUs and OpenCL all about?

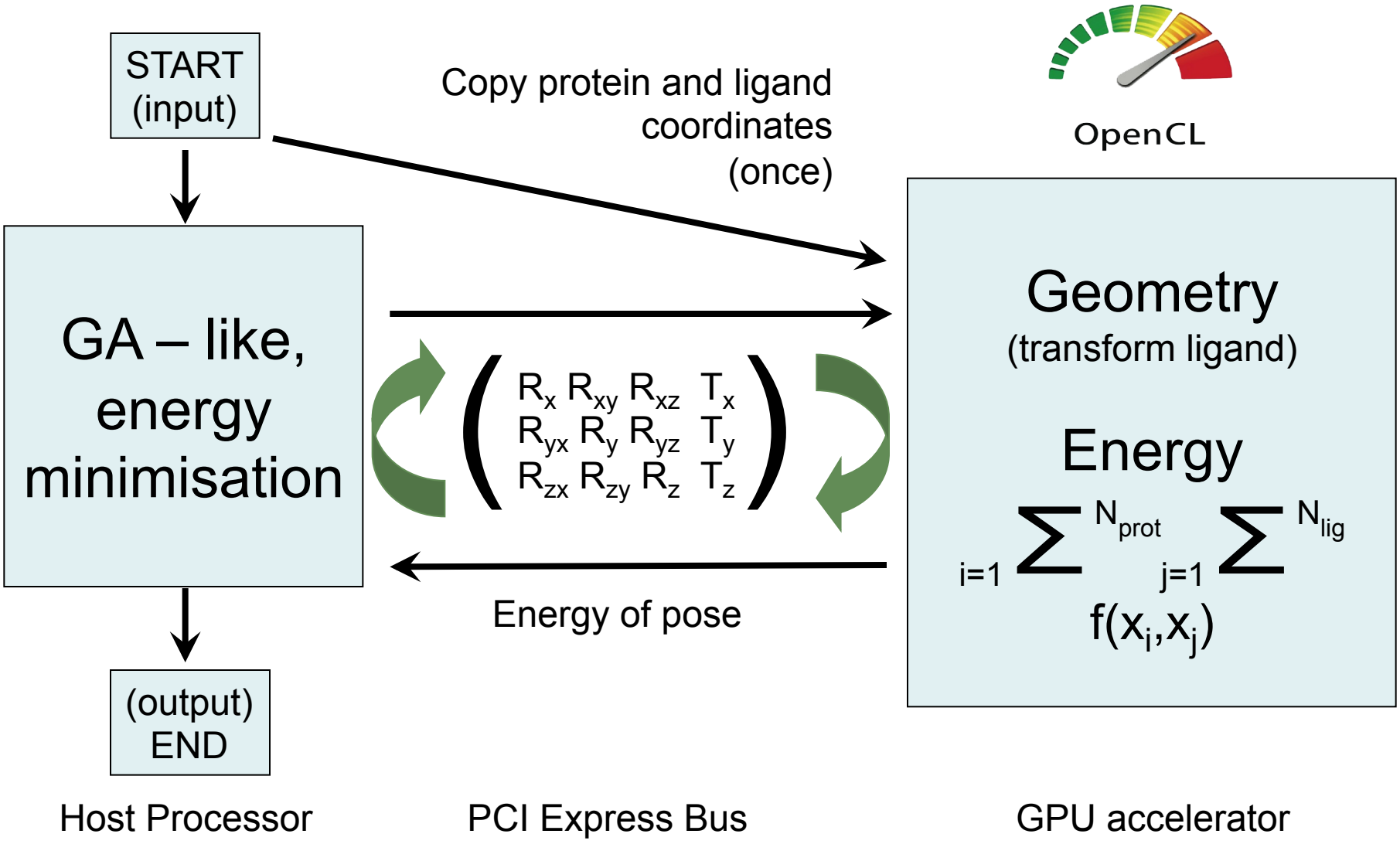
A modern computer includes:

- One or more CPUs
- One or more GPUs



OpenCL (Open Compute Language) lets programmers write a single portable program that uses ALL resources in the heterogeneous platform

BUDE Acceleration with OpenCL



"Benchmarking energy efficiency, power costs and carbon emissions on heterogeneous systems", Simon McIntosh-Smith, Terry Wilson, Amaurys Avila Ibarra, Jon Crisp and Richard B.Sessions, The Computer Journal, September 12th 2011. DOI: 10.1093/comjnl/bxr091

Systems benchmarked

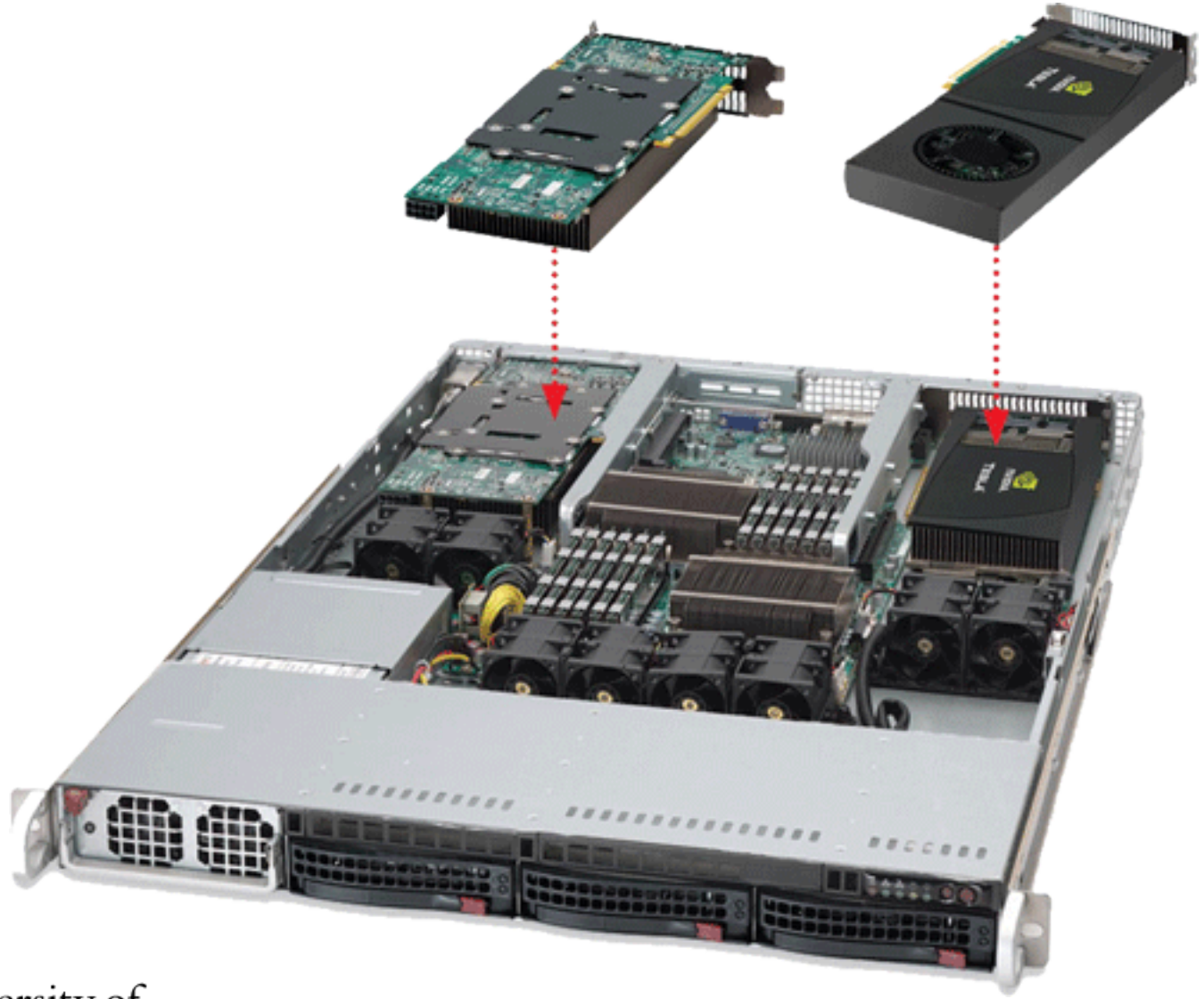
High-end:

- Supermicro 1U dual GPU server
- Two Intel 5500 series 2.4 GHz Xeon 'Nehalem' quad-core processors
- Two Nvidia C2050 'Fermi' GPUs ***or***
- Two AMD 'Cypress' FirePro V7800s

Medium-end:

- Workstation with 1 CPU & 1 GPU
- Intel E8500 3.16 GHz dual core CPU
- Previous generation Nvidia consumer-level GPU, the GTX280

Supermicro GPU server



University of BRISTOL

Systems benchmarked

Middle-end:

- Workstation based on a 3-core AMD 2.8 GHz Phenom II X3 720
- 4 GBytes of DRAM
- **No GPU!**

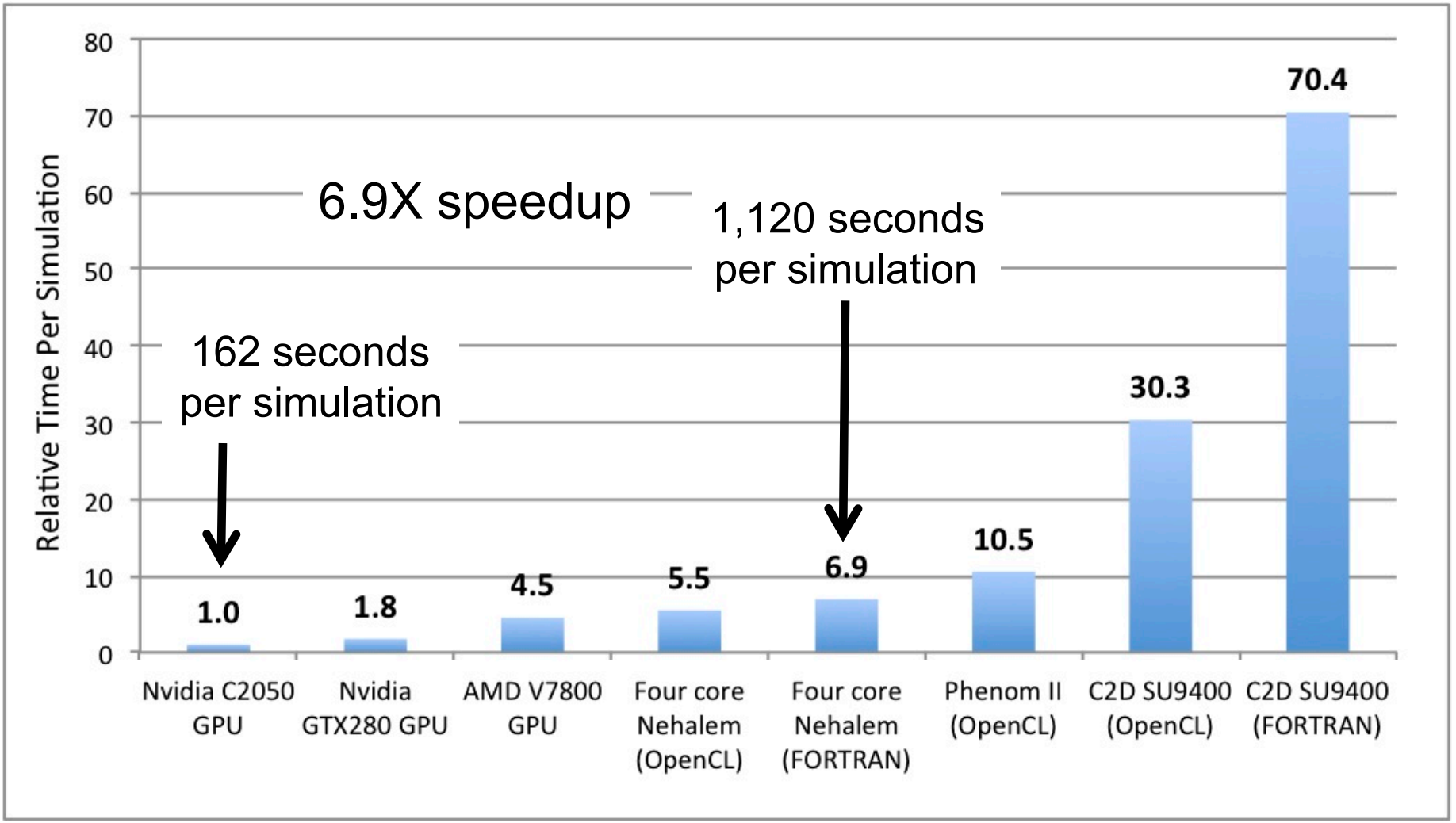
Low-end:

- Laptop based on an Intel Core2Duo SU9400 ‘Penryn’ 1.4 GHz CPU
- 4 GBytes of DRAM
- **No GPU!**

Benchmarking methodology

- Use the same power measurement equipment for all the systems under test
- Watts Up? Pro meter
- Measures complete system power ‘at the wall’
- Run as fast as possible on all available resources (i.e. all cores or all GPUs simultaneously)

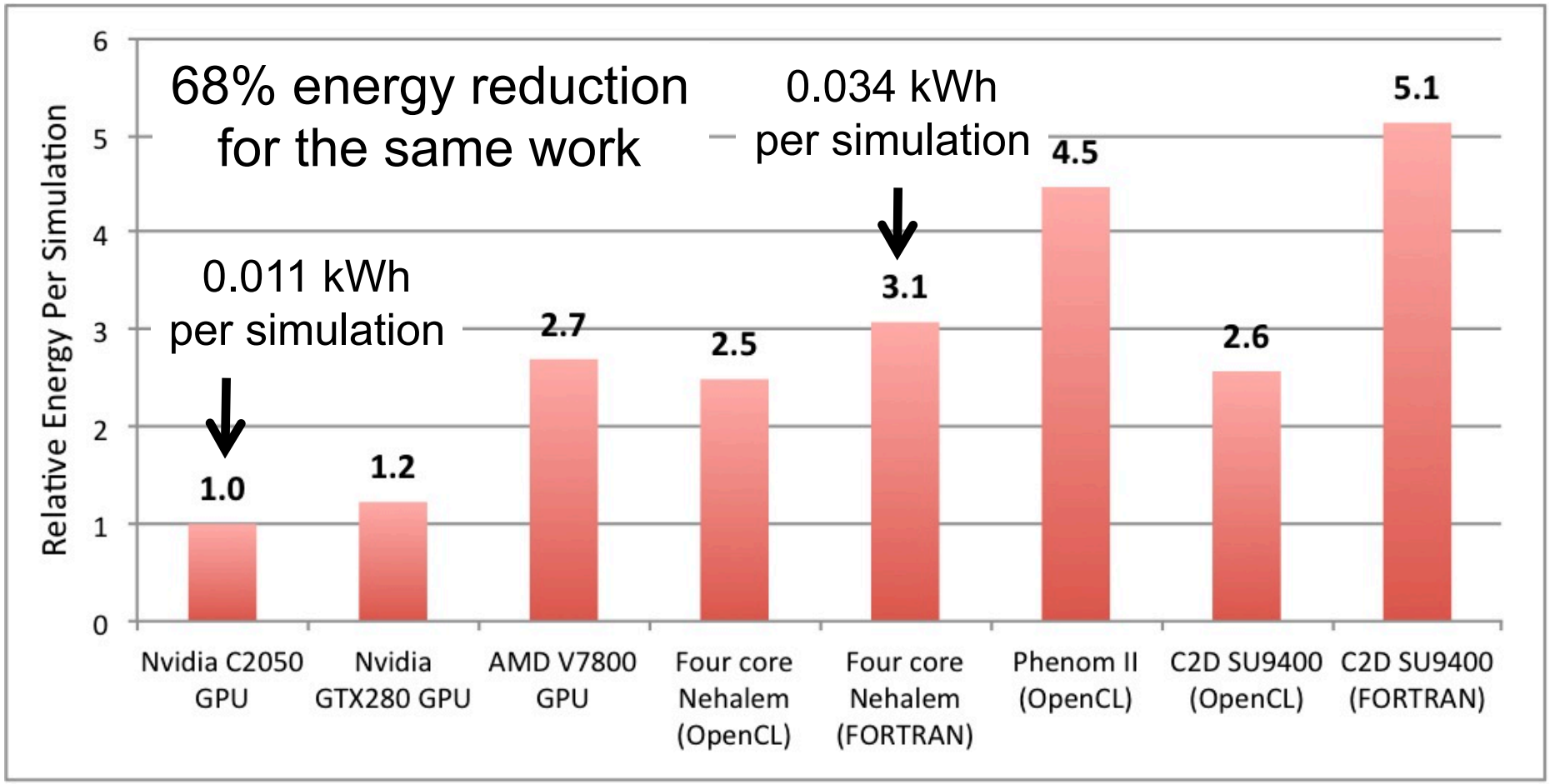
Relative performance



University of BRISTOL

Less than 2 days to screen a library of 1 million drug candidates on 1000 GPUs

Relative energy efficiency



0.011 kWh = 0.16 pence per simulation
1 million simulations → GPUs save £3,680
on energy costs for one experiment

University of BRISTOL

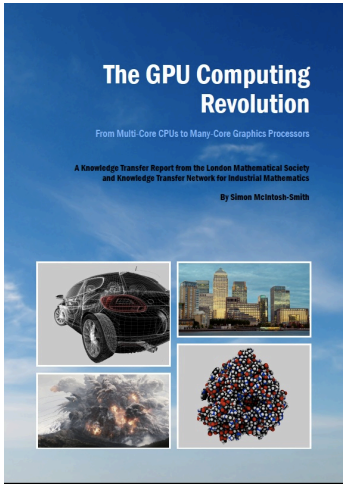
Future work

- Modify the code to use the CPUs and GPUs at the same time (*done*)
- Further optimisations (*done – code already several times faster*)
- Refine forcefield (*in progress … forever*)
- Improvements to the Genetic Algorithm and further trimming of the search space (*done*)

For an introduction to GPUs

The GPU Computing Revolution – a Knowledge Transfer Report from the London Mathematical Society and the KTN for Industrial Mathematics

- https://ktn.innovateuk.org/web/mathsktn/articles/-/blogs/the-gpu-computing-revolution



University of BRISTOL

Conclusions

- GPUs can yield big performance increases if your application has enough of the right sort of parallelism
- GPUs can can also result in significant savings in energy costs (and equipment costs)
- OpenCL can work just as well for multi-core CPUs as it does for GPUs

It's possible to screen libraries of millions of molecules against complex targets using highly accurate, computationally-expensive methods in one weekend using equipment costing << £1M

References

- S. McIntosh-Smith, T. Wilson, A.A. Ibarra, J. Crisp and R.B. Sessions, "Benchmarking energy efficiency, power costs and carbon emissions on heterogeneous systems", The Computer Journal, September 12th 2011. DOI: 10.1093/comjnl/bxr091

- N. Gibbs, A.R. Clarke & R.B. Sessions, "Ab-initio Protein Folding using Physicochemical Potentials and a Simplified Off-Lattice Model", Proteins 43:186-202,200