

Parallel Computing and Protein Design

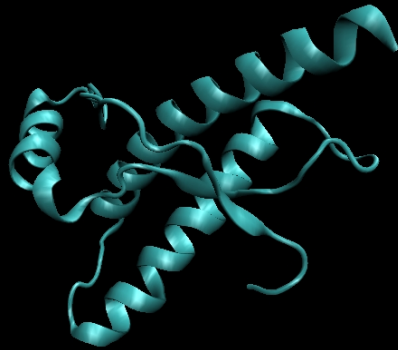
Leonor Cruzeiro¹, Paulo A. Lopes²

¹ CCMAR and FCT, Univ. do Algarve, Portugal

² CITI, FCT, Univ. Nova Lisboa, Portugal

The four protein structural classes in CATH

α



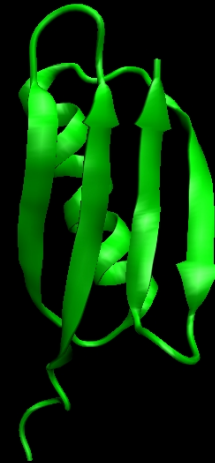
β



disordered

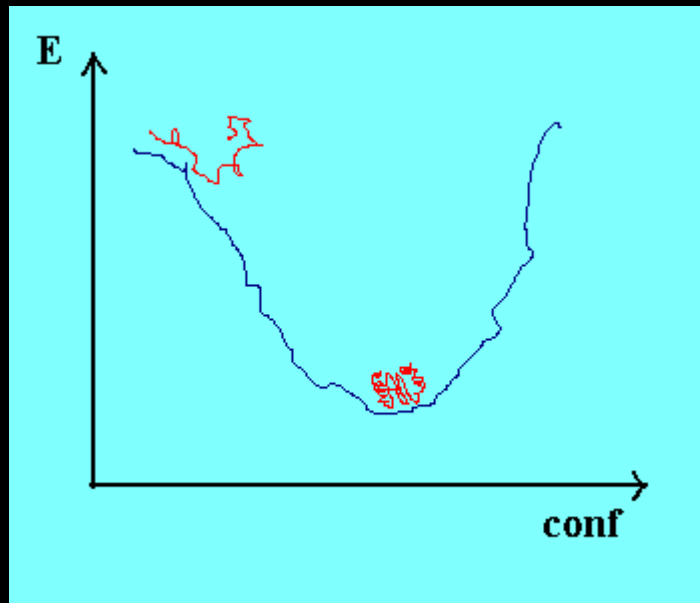


α/β

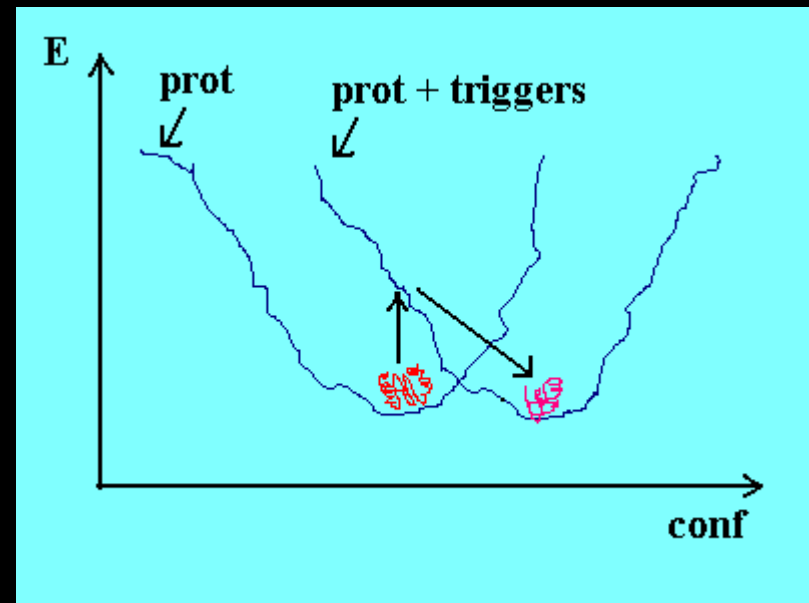


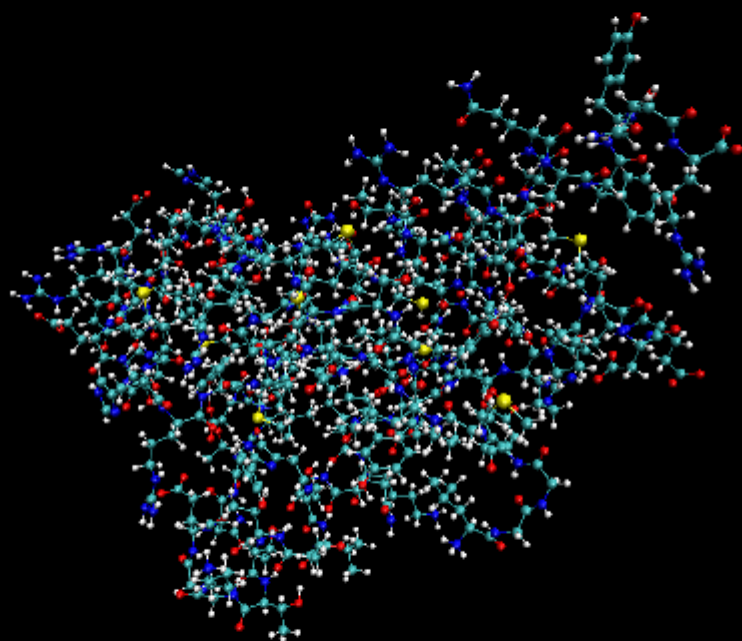
A funnel-shaped Free Energy Landscape?

Protein Folding



Conformational Change

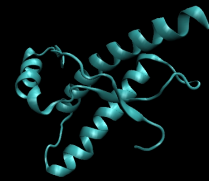




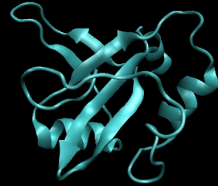
Classical Potential Energy function

$$\begin{aligned} V(\{\vec{R}_n\}) = & \sum_{\text{ligações}} K_r (r - r_{eq})^2 + \sum_{\text{ângulos}} K_\theta (\theta - \theta_{eq})^2 + \\ & + \sum_{\text{torsões}} \frac{E_n}{2} [1 + \cos(n\varphi - \gamma)] + \\ & + \sum_{i < j} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right) \end{aligned}$$

The structures of proteins ?



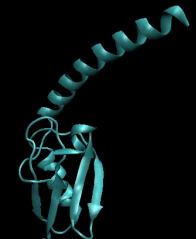
-1900



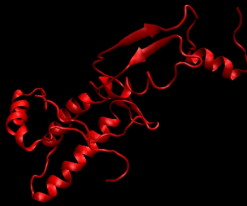
-2008



-1948



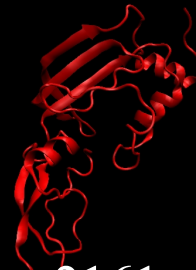
-1989



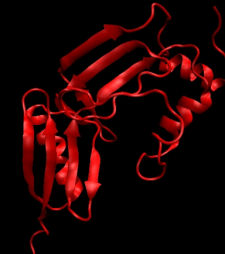
-2129



-2070



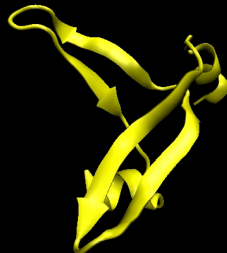
-2161



-2243



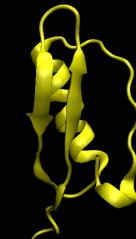
-895



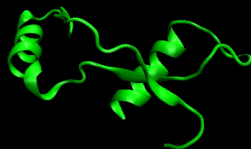
-875



-850



-1007



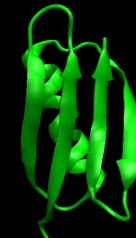
-860



-846

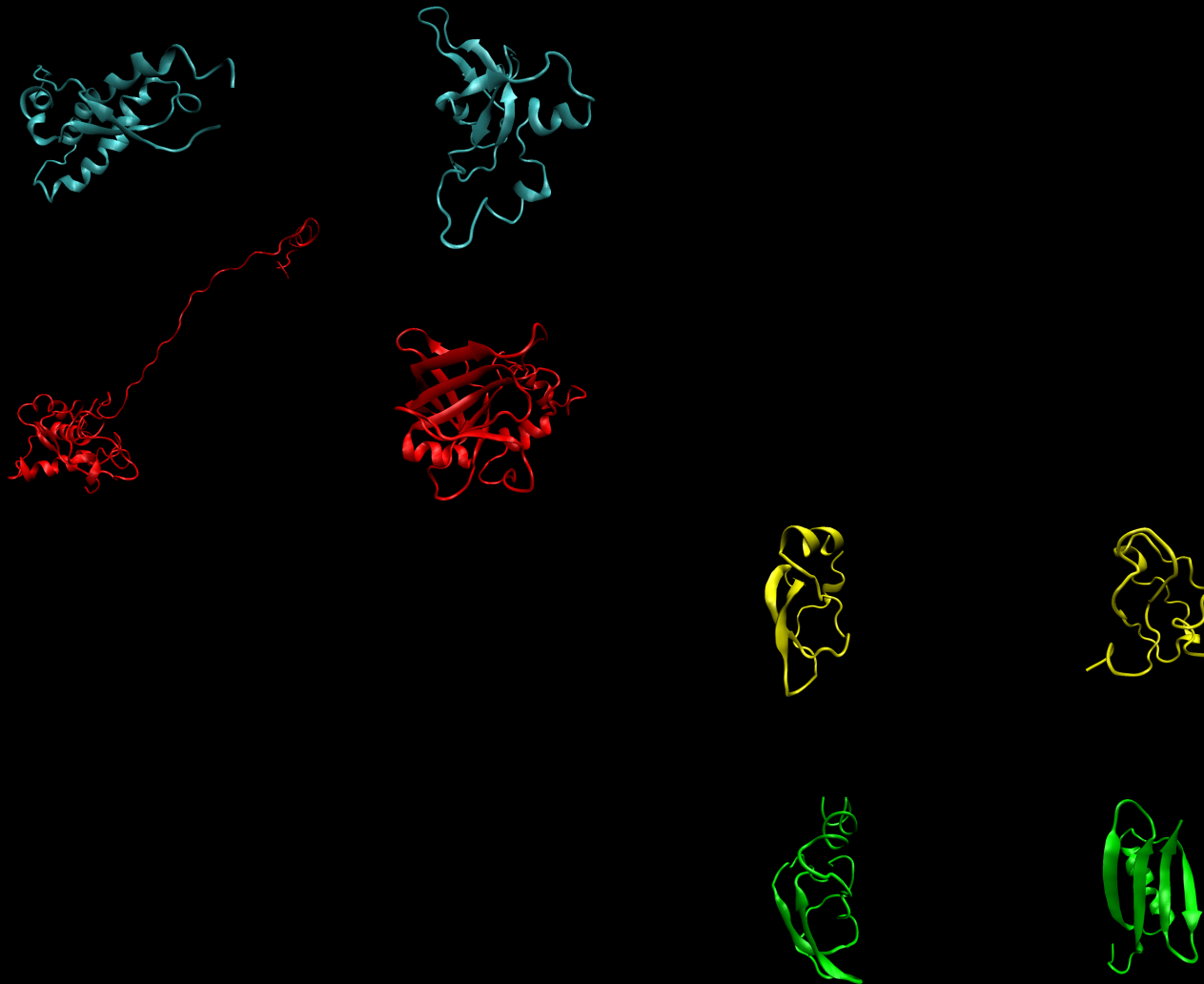


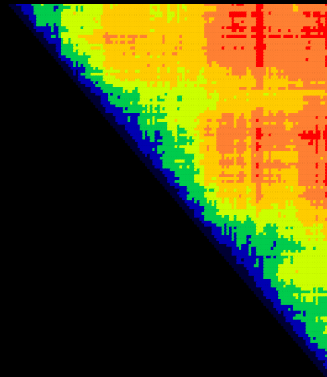
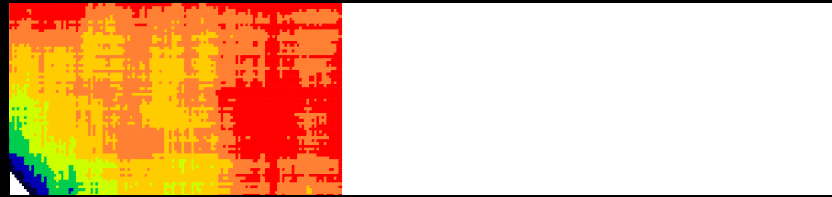
-849



-895

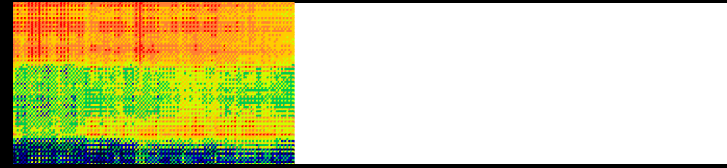
Os últimos 25 ns de uma simulação de 50 ns



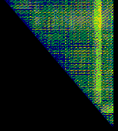
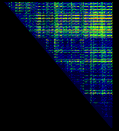


St_1QLX

St_1I0S

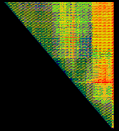


1
Q
L
X



15.1

1
I
0
S

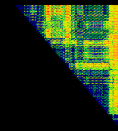
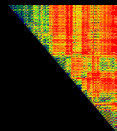


25.4

St_1AAP

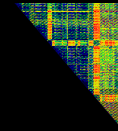
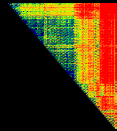
St_1IGD

1
A
A
P



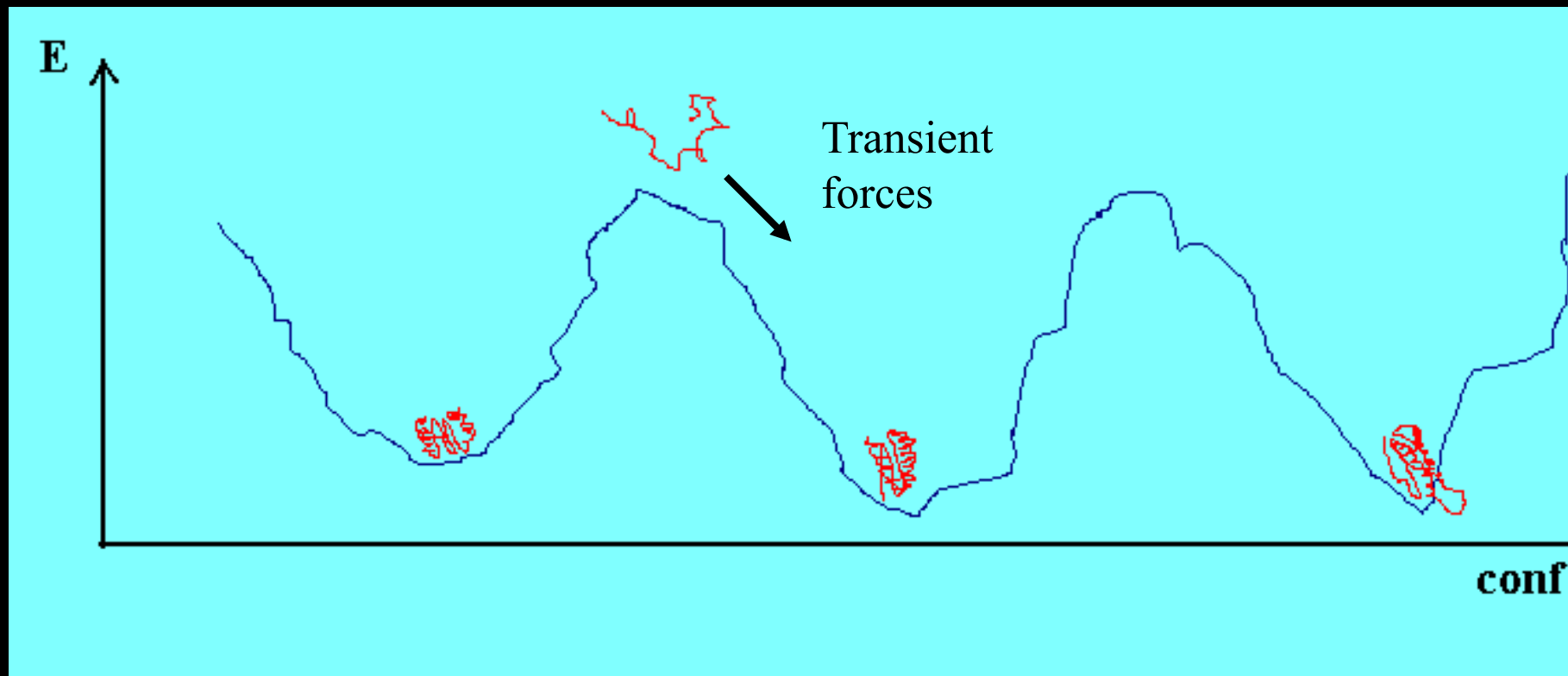
14.5

1
I
G
D



16.4

A Multi-funnel Free Energy Landscape ?



Implications for Protein Design

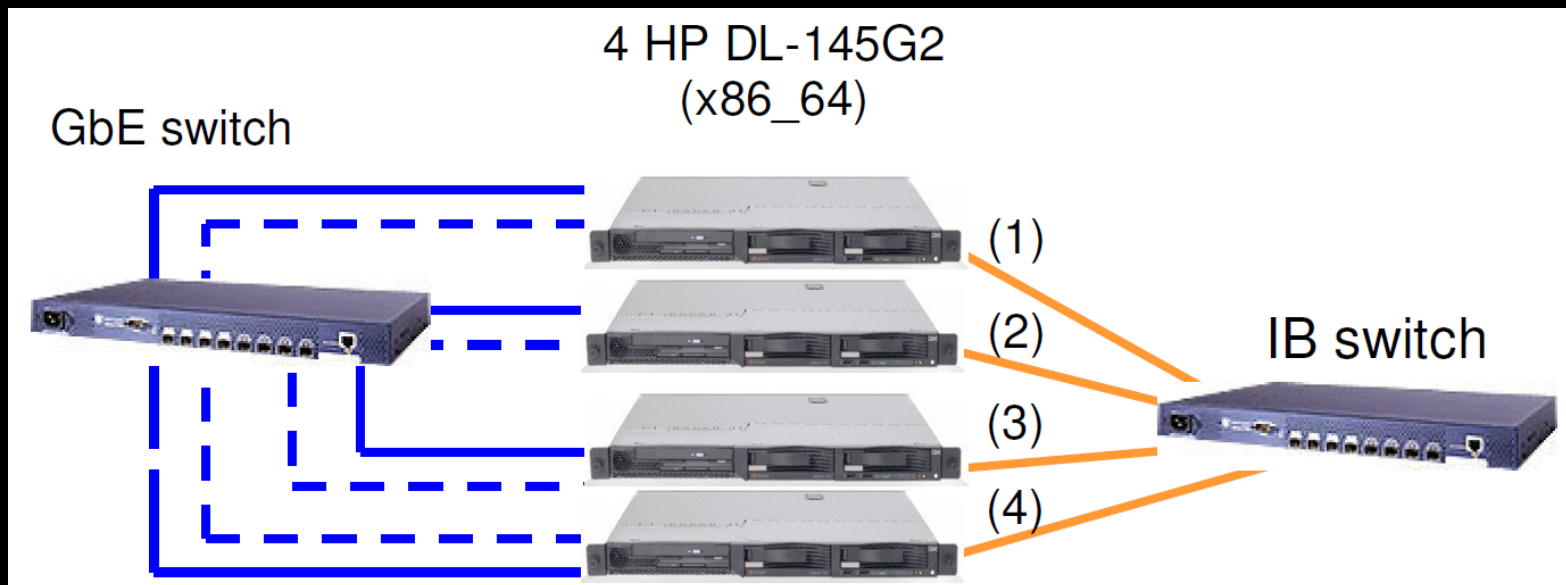
- Limitation: “only” 0.1 microsecond simulations.
- Protein Design is the inverse of Protein Folding: you want a protein to perform a specific function, that is, a protein with a specific structure, and you want to determine the amino acid sequence that will lead to that structure.
- If free energy landscape is multi-funnel shaped, then the implication is that a given amino acid sequence may well acquire different structures. Why does this not happen in cells?
- If cellular constraints on protein folding are important to define the native structure of proteins, then we must find what these constraints are if we are to make progress in Protein Design.

Acknowledgements

Partial funding from the Foundation for Science and Technology (FCT) is acknowledged.

Milipeia, Universidade de Coimbra.

The testbed



From OS bundled to the newest GCC version

Table 1. Single node performance (elapsed time in seconds).

	<i>s-core (AMD 246)</i>		<i>d-core (AMD 275)</i>	
<i>Cores</i>	<i>1</i>	<i>2</i>	<i>2</i>	<i>4</i>
<i>GCC 4.1.2</i>	<i>19996.6</i>	<i>10118.3</i>	<i>9044.8</i>	<i>4637.1</i>
<i>+ optimization</i>	<i>18855.4</i>	<i>9565.8</i>	<i>8475.8</i>	<i>4364.7</i>
<i>GCC 4.4.3</i>	<i>18721.3</i>	<i>9509.5</i>	<i>8536.9</i>	<i>4381.0</i>
<i>+ optimization</i>	<i>17369.0</i>	<i>8882.2</i>	<i>7951.4</i>	<i>4089.9</i>
<i>Optimization gains</i>	<i>15.1%</i>	<i>13.9%</i>	<i>13.8%</i>	<i>13.4%</i>
<i>Speedup</i>		<i>1.96</i>		<i>1.94</i>

Speedup as a number of cores in GbE and IB

Table 2. Cluster performance (elapsed time in seconds).

	<i>Gigabit Ethernet (GbE)</i>				<i>Infiniband (IB)</i>			
<i>Nodes</i>	2	2	2	4	2	2	2	4
<i>Cores</i>	2	4	8	12	2	4	8	12
<i>GCC 4.1.2 + Opt</i>	9550.9	5011.4	2515.5	1864.7	9517.1	4907.6	2302.7	1763.1
<i>GCC 4.4.3 + Opt</i>	8833.5	4675.5	2376.5	1817.0	8814.5	4581.7	2155.9	1667.6
<i>GCC version gains</i>	8.1%	7.2%	5.9%	2.6%	8.0%	7.1%	6.8%	5.7%
<i>Speedup (w/ GCC 4.4.3)</i>		1.89	3.72	4.86		1.92	4.09	5.29
<i>Speedup (w/ clock adjust)</i>		1.89	3.38	4.63		1.92	3.72	5.03

Speedup as a number of cores in GbE and IB

