Porting biological applications in GRID: An experience within the EUChinaGRID framework

Giovanni Minervini, Pier Luigi Luisi, Fabio Polticelli
Department of Biology, University Roma Tre, Italy
polticel@uniroma3.it

The number of natural proteins, though apparently huge, represents just a tiny fraction of the theoretically possible protein sequences. In fact, a protein chain of 60 amino acids can theoretically exist in 2060 chemically unique combinations while extimates of the total number of existing proteins vary from 109 to 1013. Thus, there exists a huge number of protein sequences that have never been exploited by biological systems. Is this because natural proteins have peculiar physico-chemical properties or else they just represent a subset of the possible protein sequences generated by the contemporary action of contingency and physico-chemical forces?

Within the EUChinaGRID project framework, we are approaching the problem by a “high throughput” approach made feasible by the use of grid techniques. A large library of random amino acid sequences not present in nature is generated and the corresponding three-dimensional structure is predicted using the ab initio protein modelling software Rosetta. Our experience in porting and running this application in grid will be described together with the results obtained so far.