Preface

Over the years, the long-standing questions on disease evolution, plausible cellular behaviors and organisms’ hidden properties have become more and more pressing. They sparked the movement of experts from an array of diverse scientific areas and stimulated them to work cooperatively with the ambitious aim of deciphering the language of nature. From the onset, the field of genomics has been built on many important discoveries, beginning with the DNA double helix structure, in the 50s; the reverse transcriptase, in the late 60s; the recombinant DNA and restriction enzymes, in the 70s and, finally, the polymerase chain reaction discovery, in the early 80s. At the time, polymerase was particularly revolutionary because it was the only method used to determine the base sequence of DNA. Thus, understanding how key components “converse” in time, in space and across multiple organizational levels was the major challenge that led scientists to the sequencing of a large eukaryotic genome, such as that of humans and to the discovery of more than 20,000 different genes and of nearly a million proteins within the cell. The transition years between the 90s and first decade of the 2000 were the time of the collection of such results; they, somehow, were mathematically formalized and summarized in artificial wired diagrams.

Soon, the idea that the achievement of the whole understanding was possible by a holistic approach rather than an atomistic one became slowly appealing. Indeed, the investigation of the assembling of cell elementary parts to form complex structures represented the main activity of a promising new research field named: Systems Biology.

Progress in this area required interdisciplinary breakthroughs and intimated link between wet and dry experimentations. Due to an enormous eager interest, synthetic formalizations (or models) grew rapidly in number and size; they became unmanageable to manual inspection and compelled biologists to reduce drastically the kind and quality of the achievable analysis procedures. If computer-driven experimentation had not been introduced almost alongside, any outcome would have been incomplete. Many automatic analysis procedures were made available to the scientific community: steady state, bifurcation, robustness analyses, model checking, simulation, etc., which represented a promising manner of providing insights for new models and of feed backing the so-called hypothesis-driven science cycle.

Dry-laboratories proliferated rapidly. To better describe their biological dynamics of interest from precise perspectives, each defined a new proprietary language. As a first result, this gave rise to an explosion of new software tools along with an overall “Babel of voices”, since each talked a different dialect. The need for a common language was becoming a “must”. Standard languages (like SBML and CellML) cropped up with the primary aim of making cooperation among software tools possible. They were able to bridge the gap between the wet- and dry-scientists by broadening the range of the available computational methods. They definitely reached their goal.

On the case of modeling, successive developments focused on how to deal with more and more complex systems. Hence, increasing evidence of the constant interplay between components of different biological systems (cross-talks) proved that isolated models are quite unrealistic and that any analysis result would be imperfect. Therefore, sets of artificial models of higher chemicals density and complexity were merged and analyzed, but without any significant results. In fact, the exponential growth of the computational power required to deal with huge state-spaces had the simple effect of hindering any of the aforementioned methodologies and of causing their ultimate flop.

One of the main limitations in managing biological models comes from the fundamental difference between evident high parallelism in biochemical reactions and sequential environments employed for the analysis of these reactions. Such limitations affect all varieties of continuous, deterministic, discrete and stochastic models; undermining the applicability of simulation techniques and the analysis of biological models.

Parallel and distributed computing intends to compensate for this lack. It relies on both the intrinsic parallelism of nature and the power of multi-processors architectures. Indeed, in real life any biological aspect is (to some extent) parallel. If a chemical transformation can occur, then it does not take place only for two molecules, but, as a principle, for all molecules. In a less coarse-grained view, hundreds of independent biological transformations take place at the same time rather than in a sequential manner. For these evidences, natural phenomena can be seen as massively parallel processes, since they occur above at least two independent levels of parallelism.
Revolutionary biology demanded revolutionary computing. Great goals have been achieved since the SPARCcenter 2000 was used to assemble the genome of *H. influenzae* in 1995. That was the last model of a computer generation limited by an architecture capable of addressing only 2 gigabytes of RAM. Modern processors are almost 200 times more powerful. The cost of storage has also dropped dramatically along the years. In 1992, one terabyte of disk-space cost one million dollars; in 2009, the cost has been reduced to near a hundred dollars. Furthermore, while in 1992 10 MB per second was the typical speed in networks, today gigabit network interfaces are very common.

Thus, the very nature of large-scale computing has changed from systems relying on one or a few powerful custom-designed processors to scalable parallel systems or farms of computer. Complex problems are now broken down into a set of smaller jobs that run concurrently on multi-processor machine nodes. Any automatic procedure is approaching an alternative parallel implementation; and several, sometimes not trivial, ad-hoc synchronization policies are coming up to support their final deployment. The desirable final aim is three-fold: making frozen processes feasible; making overall computation faster; making final analysis outcome more reliable.

The ambitious aim of HiBi09 is to establish a forum to link researchers in the areas of parallel computing and computational systems biology. Experts from around the world will present their current work, discuss profound challenges, new ideas, results, applications and their experience concerning key aspects of high performance computing in biology. The hope is to excite the researchers’ interest and to encourage them to walk together and find common and long run targets, on this path.

Tommaso Mazza  
*HiBi 2009 General Chair*

"The only true law is that which leads to freedom," Jonathan said. "There is no other."

from *Jonathan Livingston Seagull* by Richard Bach