Beilstein-Institut Molecular Informatics: Confronting Complexity, May 13th - 16th 2002, Bozen, Italy

EPILOGUE - COMPLEXITY CHALLENGES RESEARCH IN MOLECULAR INFORMATICS

GISBERT SCHNEIDER

Beilstein Professor of Cheminformatics, Johann Wolfgang Goethe-Universität,
Institut für Organische Chemie und Chemische Biologie, Marie-Curie Str. 11,
D-60439 Frankfurt, Germany.
E-Mail: gisbert.schneider@modlab.de

Received: 20th Sept. 2002 / Published: 15th May 2003

"Molecular informatics" is a scientific discipline devoted to analysing and understanding the storage, processing and distribution of information encoded by molecules and molecular interactions, coined by contemporary bio- and cheminformatics research. Although this definition of molecular informatics may not be perfect, it is comparably easy to comprehend. The term "complexity" appears more vague and difficult to define. Although most of us do have an intuitive understanding of what complexity suggests, different persons will probably give a different answer to the question what complexity actually means and implies in the context of molecular informatics. The Beilstein-Workshop *Molecular Informatics: Confronting Complexity* held in Bozen, Italy, May 13-16, 2002, brought together an international group of scientists to present their research, exchange ideas and opinions, and discuss complex systems in the light of the workshop's challenging title.

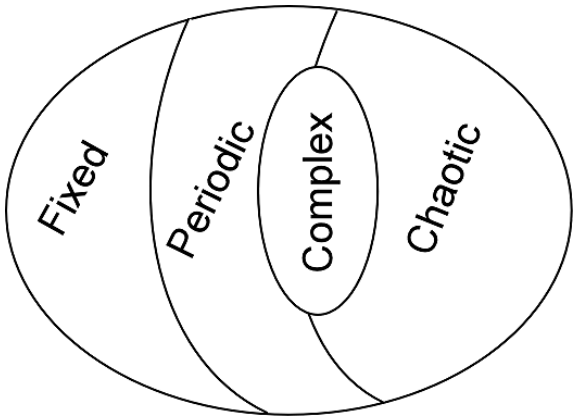

Figure 1. Complex systems may be placed between periodic and chaotic behavior. They show non-linear response, are partly unpredictable, and are characterized by the presence of noise. Graph adapted from ref. (1).

Complex systems may be characterized by three main attributes: i) partly unpredictable system behaviour, ii) non-linear response, and iii) inherent presence of noise. C. G. Langton located such systems "at the edge of chaos" (Figure 1) (1).

Typically, the objects of molecular informatics research are biological systems, like the structure and function of biological macromolecules, molecular recognition events, metabolic pathways and networks - all representing complex dynamical systems or their individual parts. It should be stressed that the term "complexity" is not the opposite of "simplicity". There are traditional scientific disciplines dealing with, e.g., algorithmic complexity addressing "orderly" systems that may be extended towards biological systems. On the other hand, a deeper understanding of biological complexity may be gained by methods such as advanced stochastic modelling and evolutionary computation, for which realizations on distributed computing facilities might be particularly well-suited (Figure 2).

The choice of methods and objects strongly depends on the scientific background and individual skills of a researcher, and several intriguing examples of both conceptual approaches are compiled in the workshop proceedings. An unifying theme during the workshop was the aim to gain insight into the behaviour of biological and molecular systems by computer simulation.

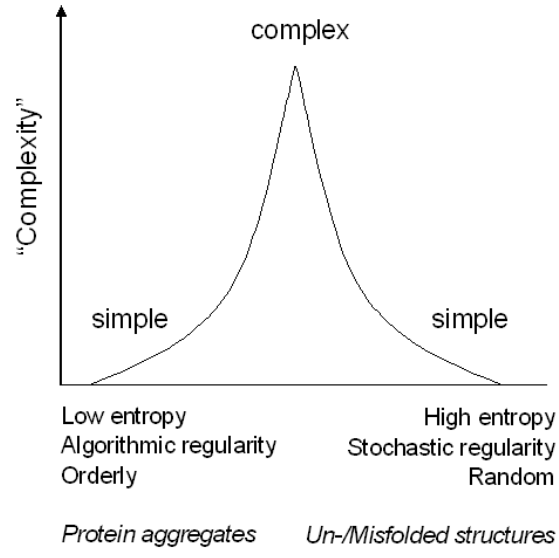

Figure 2. Complexity is not the opposite of simplicity. There are two types of problems generally regarded as tractable or "simple": orderly and random. Complex biological systems are neither entirely random or orderly, e.g. a protein's native state is neither an ordered aggregate nor unfolded. Graph adapted from ref. (6).

For example, realistic protein folding and molecular docking simulations are considered to be interrelated and represent very complex tasks. Approaches are derived from concepts abstracted

from statistical mechanics, namely, populations, and from the purely physical standpoint, binding and folding are analogous processes, with similar underlying principles (2).

According to G. P. Williams there are six ingredients to complex dynamic systems (3):

1. A large number of items ("agents")
2. Dynamism
3. Adaptiveness
4. Self-organization (i.e. order forms inevitably or spontaneously)
5. Local rules that govern each agent
6. Hierarchical progression in the evolution of rules and structures

Confronted with this list, at the end of the Beilstein-workshop the participants were asked which of the six attributes of complex dynamic systems were best covered by the lectures. The result of this non-representative survey is summarized in Figure 3, revealing a clear trend.

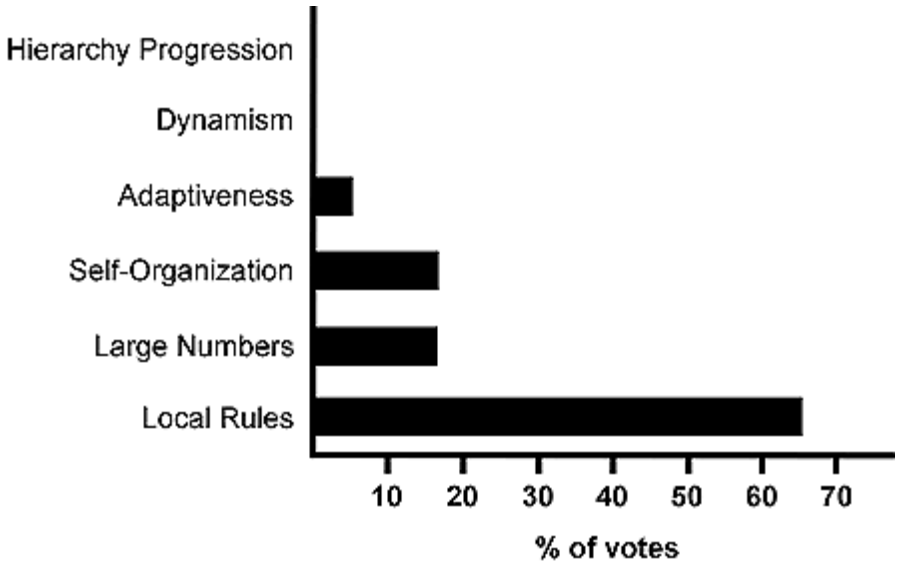

Figure 3. Result of a non-representative survey among the participants of the Beilstein-workshop 2002. The question was: "Which of the six attributes of complex dynamic systems was best covered by the lectures?"

Obviously the molecular informatics community seems to be rather familiar with the formulation of local rules and to some extent gives attention to issues related to self-optimisation and the problem of large numbers. At the same time it is obvious that essential attributes and properties of complex dynamic systems are not adequately or sufficiently treated by current molecular informatics research, namely the their dynamics, ability to adapt, and hierarchical evolution. Innovation is therefore needed to adequately treat other important attributes of complex biological systems. Generally innovation is considered to come in two

equally important guises: i) unexpected, non-linear, quantum leap innovation; and ii) linear innovation based on incremental improvements (4,5). Appropriate working environments and conditions as well as the cross-fertilization of disciplines are needed for future success. It should be appreciated that the scientific community as a whole - and the group of workshop participants in particular - forms a complex dynamic system itself. As a consequence, there is good reason for hope that system-immanent mechanisms of development and optimisation will eventually lead to progress and innovation in the most challenging areas of molecular informatics research. The current research trends identified during the workshop are functional predictions in the field of genomics and proteomics, refinement of global approaches by modular rules, development of novel representations of biological and chemical information, and adaptation of methods from engineering, computer vision, and the machine learning community.

ACKNOWLEDGEMENT

The author is grateful to the Beilstein-Institut zur Förderung der Chemischen Wissenschaften for generous support. Petra Schneider is thanked for valuable discussion during the preparation of the manuscript.

REFERENCES

[1] Langton, C. G. (1991). Life at the edge of chaos. In: *Artificial Life II* (Langton, C.G., Taylor, C., Farmer, J.D., Rasmussen, S., Eds.). Addison-Wesley, Redwood City.

[2] Halperin, I., Ma, B., Wolfson, H., Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* **47**:409-443.

[3] Williams, G. P. (1997). *Chaos Theory Tamed*. Joseph Henry Press, Washington.

[4] Austin, A. (1998). Passion versus fear as the emotion driving scientists. *Drug Discov. Today* **3**:419-422.

[5] Schmid, E. F. (2002). Should scientific innovation be managed? *Drug Discov. Today* **7**:941-944.

[6] Flake, G. W. (1999). *The Computational Beauty of Nature*. M.I.T. Press, Cambridge.