ELSEVIER

Alzheimer's & Dementia 5 (2009) 1–4

Alzheimer's & Dementia

Perspectives

In silico modeling system: A national research resource for simulation of complex brain disorders

Zaven S. Khachaturian[a,1], Joseph Lombardo[b]

[a]President, Lou Ruvo Brain Institute, Las Vegas, NV, USA
[b]Director, National Supercomputing Center, University of Nevada Las Vegas, Las Vegas, NV, USA

1. Summary

This paper presents the rationale for an initiative to create a comprehensive in silico modeling system, as a national shared resource, for simulation of complex biological systems, eg, neurodegenerative disease. The development of a well-integrated, generally-accessible computer modeling system (built upon a platform of an interlinked network of high performance/super computers -*HPC*) will enable medical researchers to translate and transform massive amounts of disconnected *information* (databases) into usable *knowledge*: an essential initial step for simulating and modeling *complex biological processes*.

The aim of this initiative is to: a) develop a blue print for action and, b) convey the vision of the next generation of research tools, technologies, and systems for the investigation of neurodegenerative disorders of later life. As a first step, a Work Group will be established to:

- evaluate the infrastructure required for creating a national ''in silico'' modeling system;
- consider a number of computer modeling approaches and databases that might be incorporated into a network of interconnected high performance computing systems;
- explore the development of a partnership among the stakeholders (eg, government agencies (NIH, NSF, DOD, DOE and their national labs), universities, pharmaceutical companies, business and finance groups, voluntary private health organizations and foundations); and
- develop a public policy blueprint to obtain the required funds, research resources and infrastructure to launch the planning and construction of the system.

The potential benefits include: a) improving the efficiency, effectiveness and speed of therapy development for neurodegenerative disorders and, b) the development of totally novel computing algorithms/devices, thinking/learning machines, robotics, and hybrid (digital/analog) computers based on a better understanding of human 'thinking' processes and brain functioning.

2. Introduction

Three decades of research have dramatically increased the prospects of developing effective treatments for neurodegenerative disorders. However, there are a number of impediments to drug-discovery and development, including: a) the philosophy, management and cost of programs, b) ethical, legal and regulatory obstacles, c) inadequate research resources and infrastructure and, d) the lack of valid models of human brain disease(s).

Perhaps the most critical barrier to therapy development is the lack of modeling system(s) that simulate the *full spectrum of clinical phenotype*(s). Although traditional modeling systems (eg, in vitro - cell cultures; in vivo - transgenic or other animal models) are available, these models do not exhibit the complete range of *biological and clinical features* of human disease. Humans with the ''disease'' represent the ultimate gold standard and the best experimental model of a ''disease;'' only humans completely express all the biological and clinical phenotypes of the disease. However, for obvious ethical reasons, people are not well-suited for manipulation of many experimental variables.

3. The Problem - Challenge

Intact animals often serve as proxies for humans with specific disease. However, animal models do not adequately represent the full spectrum of analogous biological processes (eg, precise genotypic and phenotypic traits of the human disease). Although transgenic animal models (with the human gene of the disease) have succeeded in producing some analogs of the neuropathologic and behavioral elements of Alzheimer's disease (AD), they still lack the human context with the full expression of all the relevant clinical symptoms.

[1]Corresponding author. Tel.: 702-263-9797; Fax: 702-260-9797.
E-mail address: zkhachaturian@keepmemoryalive.org

1552-5260/09/$ – see front matter
doi:10.1016/j.jalz.2008.11.003

The limitations of transgenic animal models do not represent a serious problem for many experimental settings, particularly in the early stages drug discovery (eg, receptor-binding assays or identification of potential lead molecules). The serious limitations of these models become apparent in the later stages of therapy development, because of their failure to embody the ''whole integrated physiological system''. These models are inadequate in simulating complex biologic process (eg, clinical expression of human disease) involving *nonlinear* interactions among several systems (eg, genes, molecular changes, neural systems and behavior).

One of the impediments to effective therapy development for neurodegenerative disease is the lack of knowledge about the precise functional relationships between the clinical (eg, symptoms manifested) and biological phenotypes (eg, neurobiological markers). As a result, the discovery of a biological marker for a disease or the identification of molecular target often does not translate into an effective therapy. The ill-defined or nonlinear relationship between molecular events and clinical symptoms is a major challenge for therapy development.

The difficulty of linking the biochemical and molecular events or changes at the cellular level to the functioning of the whole brain (eg, behavior or clinical symptoms of the diseases) results from the lack of conceptual models of emergent behaviors in complex systems. The clinical symptoms of dementia, human cognition, tertiary structure of proteins, functions of proteins, and the flight patterns of flocks of birds are some examples of *emergent behaviors* in complex systems. Emergent behaviors are typically context-dependent. The problem in creating an integrated picture of the relationships between molecular data (genotype) and complex behaviors (phenotype) is not unique to neurobiology or AD.

It is a challenge for many fields of study struggling to understand the behavior of a complex system through detailed analysis of the simpler (eg, molecular) constituent parts. For example, in the field of protein chemistry, detailed information on genes, amino acid sequences, or the secondary or tertiary structure of a protein do not enable the prediction of function (eg, protein-protein interactions). Thus, in the current environment with the available research tools, it is not likely that detailed information on molecular sequence-structure-function will yield the answer to the clinical issues of dementia. For example, the gene for Huntington disease has been known for nearly quarter of century, yet this knowledge has not produced any effective therapies.

To maximize the yield from the enormously rich and detailed molecular information (data) on human brain disorders, a holistic modeling approach is needed. The proposed modeling system, based on the HPC platform, will address this challenge by linking and integrating information from various hierarchies of organization or different levels of complexities, eg, regulation of gene expression, metabolic pathways, signaling cascades, and neural systems/behavior.

4. In Silico Models

An in vivo experiment allows the investigator to study a phenomenon in the full biological context of a living organism (eg, electrical activity of a single nerve cell in a living brain). In contrast, in an in vitro experiment, the same phenomenon is studied in an isolated nerve cell removed from the living organism; an isolated biological process that has lost its living context. In studies of the same phenomenon conducted in silico,[2] the investigator attempts to *abstract and isolate the logical form* of the natural biological processes from their material substrate. In vitro models are often used in investigations of complex biological phenomena because it is easy to prepare; to control the system's boundary conditions; to hold many variables constant; to manipulate selected variables and to observe their effect on behavior. In vitro experiments are often preferred to in vivo experiments in cases where the biological context in which the process is embedded is so complex that it is difficult or impossible to manipulate the process precisely enough to ensure repeatability. The downside of in vitro experiments is precisely this loss of the complex context within which the process was embedded, quite possibly causing the observed in vitro and in vivo behaviors to be very different from each other.

In contrast, in silico experiments allow precise observation and control of experimental conditions while modeling the ''process'' of interest in the full biological context. Thus, once an investigator fully understands the behavior of a biological process, by precise manipulation and experimentation in vitro, a ''model'' of that process can be created. If an investigator can abstract models for all of the components involved in a complex biological system, it will be possible to synthesize a model of the whole complex by linking together the various models (or levels of complexities in a hierarchically organized system) – outputs of one to inputs of others - and iterating them forward in each other's presence in a computer. This approach is particularly useful in complex systems with nonlinear interactions among its constituent components, eg, genes-sequence-structure-function-behavior continuum. In silico modeling enables the investigator to directly observe the manner in which the behavior of the whole system emerges from its parts. Because this takes place in a computer, the investigator can prepare and manipulate the system in precisely repeatable ways, and observe every aspect of the system without interference. Thus, in silico modeling could provide the full biological context of in vivo experiments without giving up the precise experimental control needed for repeatability that is provided by in vitro experiments.

The emerging in silico modeling systems, particularly the class of modeling generally referred to as ''multi-modeling'' approaches, could add an important new tool for investigators in understanding the relationship between molecular

[2] The term ''in silico'' is analogous to the terms ''in vivo'' and ''in vitro.'' It means ''in the chip,'' or ''in the computer.''

phenomenon in the neuron and the emergent behavior as clinical symptoms of the whole complex system. These technologies do not assume any single approach; rather they provide the conceptual framework for integrating a wide-spectrum of modeling approaches, often spanning many different formalisms, spatial and/or temporal scales and resolutions.

A multi-modeling framework can manage the integration of all different component-specific approaches, so that the investigator can study not only the critical interactions among all components but also the resulting emergent dynamics of the whole system. Multi-modeling technologies are essential in modeling complex systems, eg, clinical symptoms of AD, which cannot be solved by the traditional analytic approach. The traditional methods rely on breaking up the system into its constituent processes and attempting to study, and/or solve, each constituent part of the complex system independently. The performance of most complex systems (eg, neural networks, signal transduction paths, protein-protein interactions) must be studied as whole systems. The behavior(s) of complex systems are the emergent characteristics resulting from the particular interactions between their constituent parts or processes, not within those constituent parts or processes.

5. The Need – 'Thinking Tool'/ Neuroscience Information Management System

The dramatic advances in neurobiology during the last three decades were largely the result of developments in analytical tools (eg, PCR) and the availability of shared research resources. The need for an in silico modeling system, as a new *research tool*, is more urgent today then ever, if a quantum leap in knowledge on neurodegeneration is to be achieved. The proposed modeling system should be created as a shared research resource built on a national network of well-integrated databases. The establishment of such a *neuroscience information management system* will create a highly integrated user-friendly tool to simulate or generate hypotheses and will address the limitations of in vivo and in vitro experimental platforms. Fortunately, the effort to establish such a critical research infrastructure will not be difficult because several in silico modeling approaches, databases and other important components are already in place. There is a growing community of scientists with the interest and expertise in computational biology/neuroscience that, in partnership with computer design and software engineers, could help define the specifications for such a system.

The *neuroscience information management system* will provide a shared high-performance computing research infrastructure based on ''grid'' and ''grid cluster'' technology. High-performance computing generally implies a system (or array) of machines capable of storing large data sets and performing calculations at a high rate. The grid is an emerging infrastructure that will fundamentally change the way we think – and use – computing. The word ''grid'' was chosen by analogy from the electrical power grid, providing pervasive, dependable, consistent, and inexpensive access to advance resources [1]. The term ''the grid'' was coined in the mid-1990s to denote a proposed distributed computing infrastructure for advanced science and engineering [2]. The goal was to create the illusion of a simple yet large and powerful self-managing virtual computer out of a large collection of connected heterogeneous systems sharing various combinations of resources. Eventually, users will be unaware they are using any computer but the one on their desk, because it will have the capability to reach out across the national network and obtain whatever computational resources are necessary [3].

Grid technology allows organizations to use numerous computers to solve problems by sharing computing resources. The systems tied together by a grid might be in the same room or distributed across the globe; running on multiple hardware platforms; running different operating systems; and owned by different organizations. Grid users experience, essentially, a very large virtual computer with a potential of accomplishing useful work. Grids have the potential to deliver an increase in performance of three orders of magnitude within five years, and five orders of magnitude within a decade [2].

The benefits of utilizing grid computing technology are as follows:

- Computing Resource Aggregation: allows users to treat geographically dispersed systems as one virtual computer with efficient resource management. Users are able to form virtual organizations that collaborate on common problems. This enables the sharing of applications and data over the Internet, much as if they were sharing a single (virtual) computer.
- Database Sharing: allows access to any remote database within a grid.
- Collaboration: allows widely dispersed organizations to work together on a project, sharing everything from valuable data to software applications. This provides an opportunity to lower the total cost of computing by enabling the sharing, efficient optimization, and overall management of grid computing resources over the Internet.
- Computational Strength: allows the execution of large problems demanding huge computing resources by enabling the aggregation of computing power and access to storage and other resources over the Internet.

Innovations in computing algorithms promise an unprecedented opportunity to simulate or model complex 'emergent behaviors,' eg, clinical symptoms such as memory disorders. Moreover, recent advances are converging to make it timely and feasible to formulate a national strategic plan for constructing in silico modeling system(s) for human neurodegenerative disorders. The availability of super-computers has made it possible to develop large and diversified databases for computer abstractions of neural function/dysfunction. These technologies should enable the creation or

construction of prototypes for use in complex-systems modeling environments. The development of such a system will enable investigators to address the limitations of in vivo and in vitro experimental platforms in therapy development. The feasibility of these techniques has been established in other industries, including aerospace, oil and gas, automotive, and semiconductors. Multi-modeling systems have been widely used in other areas of science to study emergent behaviors of complex systems such as the impact of atmospheric changes on weather and the behavior of commodities market traders. Adoption of such approaches should increase the understanding of the complexities of genetic mutations, molecular changes, aberrations in signal transduction pathways, abnormalities in neural networks and their translation into clinical symptoms. Such knowledge, which does not exist now, will significantly improve the efficiency and pace of the drug discovery process. Additionally, research will continue over the next few years on modifying algorithms and codes to take advantage of the availability of the increasing number of parallel computing systems, clusters, and grid networks.

6. Conclusions

The creation of a comprehensive *neuroscience information management system* as a platform for in silico modeling systems/approaches will provide a unique research tool in drug discovery and development research. The enormous volumes of data and the wealth of isolated information bits often cannot be used by a scientist unless dynamic systems models integrate the data and convert the information into knowledge. In silico modeling will:

a) Reduce the cost and increase the efficacy of drug discovery and physiological research,
b) Improve the rationale for particular target and lead molecule selection,
c) Permit simulation studies or allow ''what if'' testing of new ideas and
d) Improve the value derived from large research investments.
e) Extend access to specialized, high-cost health care resources.

References

[1] Foster I, Kesselman C, editors. The Grid: Blueprint for a New Computing Infrastructure. San Francisco, CA: Morgan Kaufmann Publishers, Inc.; 1999.
[2] Casanova H, Dongarra J. Netsolve: a network server for solving computational science problems, Tech. report CS-95-313, University of Tennessee; November 1995.
[3] Catlett C, Smarr L. Metacomputing. Communications of the ACM 1992; 35:44–52.