ELSEVIER

Alzheimer's & Dementia ■ (2017) 1-11

Alzheimer's & Dementia

Perspective

Knowledge-driven computational modeling in Alzheimer's research: Current state and future trends

Hugo Geerts[a,b,*], Martin Hofmann-Apitius[c], Thomas J. Anastasio[d], on behalf of the Brain Health Modeling Initiative

[a]*In Silico Biosciences, Berwyn, PA, USA*
[b]*Perelman School of Medicine, Univ. of Pennsylvania*
[c]*Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, Germany*
[d]*Department of Molecular and Integrative Physiology, and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA*

Abstract Neurodegenerative diseases such as Alzheimer's disease (AD) follow a slowly progressing dysfunctional trajectory, with a large presymptomatic component and many comorbidities. Using preclinical models and large-scale omics studies ranging from genetics to imaging, a large number of processes that might be involved in AD pathology at different stages and levels have been identified. The sheer number of putative hypotheses makes it almost impossible to estimate their contribution to the clinical outcome and to develop a comprehensive view on the pathological processes driving the clinical phenotype. Traditionally, bioinformatics approaches have provided correlations and associations between processes and phenotypes. Focusing on causality, a new breed of advanced and more quantitative modeling approaches that use formalized domain expertise offer new opportunities to integrate these different modalities and outline possible paths toward new therapeutic interventions. This article reviews three different computational approaches and their possible complementarities. Process algebras, implemented using declarative programming languages such as Maude, facilitate simulation and analysis of complicated biological processes on a comprehensive but coarse-grained level. A model-driven Integration of Data and Knowledge, based on the OpenBEL platform and using reverse causative reasoning and network jump analysis, can generate mechanistic knowledge and a new, mechanism-based taxonomy of disease. Finally, Quantitative Systems Pharmacology is based on formalized implementation of domain expertise in a more fine-grained, mechanism-driven, quantitative, and predictive humanized computer model. We propose a strategy to combine the strengths of these individual approaches for developing powerful modeling methodologies that can provide actionable knowledge for rational development of preventive and therapeutic interventions. Development of these computational approaches is likely to be required for further progress in understanding and treating AD.

Keywords: ■■■

1. Introduction

Neurodegenerative diseases such as dementia are complex multifactorial pathological processes with long presymptomatic incubation periods. Longitudinal cohort studies, such as ADNI [1] and the EPAD [2], have begun deep phenotyping of the progression of Alzheimer's disease (AD). In addition, a large number of preclinical animal studies together with multidimensional clinical observations have identified over 1000 different factors that are proposed to contribute to AD pathology [3,4]. Together with the

*Corresponding author. Tel.: +1-267-679-8090; Fax: ■■■.
E-mail address: Hugo-Geerts@In-Silico-Biosciences.com

http://dx.doi.org/10.1016/j.jalz.2017.08.011
1552-5260/

significant variability of longitudinal patient trajectories [5] and the limited clinical translatability from animal models [6], understanding the causal biology of AD poses a real challenge. But it is a challenge that must be met because a thorough understanding of biology is an essential contributor to the success of any biomedical R&D project [7] and the development of AD therapeutics is no exception.

Although it is still arguably incomplete, the sheer magnitude of the existing body of information on AD pathophysiology itself poses a barrier to understanding. Even when presented with powerful visualization tools, the nature of the interactions between the many factors that are supposed to underlie AD is impossible for the unaided human brain to discern. With the capacity of human working memory (the memory required for cognitive manipulation of information) estimated to be about four items [8], it is easy to see why researchers rely on intuition to decide the direction of their AD research. The impressive size of the existing AD data set is testament to the effectiveness of intuition as a guide for the discovery of additional AD facts, but our failure so far to integrate this information for the development of effective AD treatment equally points to the limitations of intuition. Indeed, intuition has often led the AD field astray.

To take the most salient example, the identification of mutations in the amyloid precursor protein (APP) gene led to the formulation of the amyloid-β (Aβ) hypothesis [9]; however, the biology of amyloid processing and impact on functional outcomes are much more complex than had been anticipated [10–12], and the extrapolation of insights from early-onset and familial AD to sporadic late-onset AD might be invalid [13]. Therapeutic strategies aimed specifically at Aβ modulation so far have yielded discouraging results in clinical trials, and the literature is now replete with critiques of the amyloid hypothesis. Although Aβ most likely plays an important role in AD pathophysiology, the importance of other factors in the progression of this disease is now without doubt. The major challenge has become one of understanding the interactions among the numerous factors that underlie AD.

These interactions play out at many different scales, from intracellular processes to synaptic functions and from neuronal connections and differential activation of brain regions to behavioral outcomes that are often difficult to quantify. The AD pathological process is not only multifactorial but also multilevel, and the available data vary greatly in the precision in which they are reported. Given the restricted information capacity of the human brain, it is fair to say that we will not be able to understand these multifactorial, multilevel processes without computer assistance in the form of analysis, modeling, and simulation.

In life sciences, sophisticated bioinformatics approaches, including those involving deep learning, have resulted in the identification of many associations with clinical phenotypes. As an application, AlzBase is a curated database on gene dysregulation in AD derived from data in various omics fields [14]. Similar approaches have merged human-curated gene or protein databases such as Online Mendelian Inheritance in Man to find candidate gene mutations that could contribute to AD [15–17]. Although they have been very important in identifying key processes, most of these correlative approaches lack quantitative aspects that are absolutely necessary in pharma R&D and public health decisions. The well-known *cum hoc ergo propter hoc* logical fallacy means that correlation does not imply causation; although observed associations have provided valuable clues into AD pathogenesis, we would not necessarily expect them to point directly to effective interventions. The limitations of correlative approaches have been described in recent critiques of genome-wide association studies in various contexts [18,19], and a recent review expresses skepticism of massive omics-based studies in AD and argues for better understanding of the actual interactions that occur within and between different cell types [20]. Concerns about correlative studies leave opportunities open for other approaches.

Meanwhile, an entire domain of knowledge exists on AD, which has largely been ignored by computational modelers and analysts. Decades of basic and clinical research have provided a wealth of data on the underlying causes of AD and on interventions that alter its course. Data on cause/effect relationships that have been well established experimentally in the AD field can be used to generate quantitative predictions and to complement correlational knowledge.

We argue that for understanding AD pathogenesis, computational models are needed that simulate pathogenic processes for generating cause-and-effect predictions. Computational AD models exist. Most are not multilevel but focused on a single level and involve relatively few factors. Some use statistical physics methods or molecular dynamics simulations to understand phenomena such as Aβ agglomeration or interaction of Aβ with key proteins [21–28]. Related but more pharmacologically oriented studies use computational chemistry and cheminformatics to search for quantitative structure-activity relationships to identify, for example, inhibitors of key enzymes in the formation of Aβ [29,30]. Other computational studies use biomimetic models of neural networks with differing degrees of biological detail to assess, for example, the vulnerability of specific brain sites to neurodegeneration or the benefits of potential AD treatments [e.g. [31–33]]. Some of the earliest computational models of AD pathogenesis were connectionist-style neural network models of language and other cognitive impairments [34–36].

In this report, we will review three different computational approaches that attempt to integrate various information modalities into dynamic and actionable platforms with a strategy on how they could be combined. Integrative modeling of data and knowledge aims to bridge data-driven and hypothesis-driven approaches by combining unbiased mining of databases with the scientific literature, attempting to quantitatively identify mechanisms that play a role in AD pathology and to