

A Top Twelve List for Biocomputing: A Decade of Progress and Challenges Ahead

By Ruth Nussinov, PhD

Editor's Note:

In addition to asking 10 experts to weigh in on Eric Jakobsson's 2005 Top Ten Challenges for the field of biomedical computing (Top Ten Retrospective, in this issue), *BCR* asked **Ruth Nussinov, PhD**, to reflect on the last decade and offer a new list. Nussinov, who is a professor of genetics at Tel Aviv University School of Medicine and senior principal scientist and principal investigator at the National Cancer Institute, came up with a Top Twelve that melds advances of the last decade with related challenges for the future. Not surprisingly, this list overlaps somewhat with Jakobsson's. But the two stories make a nice pairing: Their similarities clarify the continuity in the field, while their differences let the shifting nature of the scientific enterprise shine through. >

W

When asked to choose the most significant computational biology advances of the last decade, I welcomed the opportunity to reflect on the field. But soon the daunting nature of the task became clear: The field has come a long way in ten years and is so broad (covering the entire range of biomedicine) that the possible choices are numerous. Moreover, there may well be entire categories of research that haven't come to my attention. On top of that, there's the question of subjectivity: The significance of any research advance is a matter of opinion.

To make the job tractable, I opted not to identify specific research papers but instead to focus on broad topic areas where computation has made, and will continue to make, a major contribution.

It is my sense that, these days, computational biology is closely tied to experimental research. And this is an advance in itself (see #1). Thus, the progress described here is not purely computational in nature; it is tied to biomedicine. And that's as it should be.



1. Mainstreaming Computational Biology

Ten years ago, as the National Institutes of Health were preparing to fund the National Centers for Biomedical Computing, we lived in a different world. The field of computational biology, though established, was dispersed and not entirely trusted by mainstream biological and medical researchers. Today, by contrast, computational biology is intimately connected to the rest of biomedicine. It's easier to collaborate across disciplines. And laboratory researchers have a better understanding of the value of using computational models for hypothesis generation, as well as the need to iterate through a cycle of modeling and laboratory testing. In sum, computational biology has become more closely integrated with experimental work, to the betterment of biomedicine as a whole.

It's a change that's also reflected in societal expectations about computation's potential generally. This attitude shift is felt throughout the field: Computational biologists now get the respect they are due.

Looking to the future, I see computational biology increasingly taking the lead in the medical sciences. I expect experimentalists will frequently find themselves testing hypotheses generated by computational analyses of massive datasets.

2. Individualizing Sequencing

Over the past decade, sequencing technologies have changed the face of genomics at an unprecedented pace, spurring new opportunities in computational biology. The one billion dollar cost of the first two human genome projects has dropped by a factor of a million. Soon a completely different and novel approach called single molecule sequencing may become viable. The method involves pulling a DNA strand through a membrane with a nanopore designed to measure tiny base-specific changes in impedance across the pore.

This technique has great promise. For example, it will be possible to compare corresponding sequences in diseased and healthy tissue of a single human being. Such methods may lead to early screening of ailments and epidemics, and suggest more effective personalized treatment decisions.

Eventually, single molecule sequencing may also permit delivery of drugs (attached to DNA via sticky ends) to specific locations using complementary sequences. Such an approach has already been proposed and could be used to target malignant genomic aberrations.

3. Imaging Molecules in Action

A vast improvement in our ability to image living systems at all levels has been the key to many crucial developments and is likely to remain so into the future. Over the last decade, X-ray crystallographic imaging has proven its efficacy by revealing structures of thousands of proteins and by successfully decoding the structure of a whole RNA-protein complex—the ribosome.

Recent advances in nuclear magnetic resonance (NMR) imaging have led to gains of a complementary but more dynamic nature. Whereas X-ray crystallography captures the single dominant native structure of a protein, NMR can often verify the existence of several and even dozens of transient alternate forms of a protein structure. Understanding the range of possible shapes a protein can take will help

in the development of drugs.

There have also been remarkable advances in the imaging of single cells, individual organelles, and even single molecules using fluorescence, electron microscopy, atomic probes and other techniques. These allow researchers to follow morphological and dynamical changes over reasonably short timescales.

In sum, molecular imaging tools now at our disposal not only allow the reconstruction of static protein structures, they can track transitions between structures, as well as reveal the kinetics and dynamics of such things as gene transcription and splicing. As we move forward, such tools, combined with computational approaches, may be able to track processes in the living cell across space, time and environment, ultimately revealing cross-scale relationships such as those between cellular outcomes and morphogens (signaling molecules involved in tissue development), genetic mutations, post-translational modifications or pathogenic proteins.

4. Going Beyond the Genome

In the nature versus nurture debate, there's been a surprising shift toward nurture over the last ten years, with computational analyses playing a role in discoveries on both sides.

In support of nature, an ever-increasing number of human traits—including even the tendency for happiness—have been found to have a genetic component. And epigenetic activity affecting the dynamics of gene expression has been found in the areas of DNA that don't code for specific proteins (areas previously considered “junk DNA”). Indeed sections of DNA that are physically far away from a protein-coding location can critically affect function.

On the side of nurture, recent developments show the importance of DNA methylation and post-translational modifications of proteins as well as other changes that aren't built into the genome but instead develop during the course of an organism's life. For example, prion disease and type A diabetes directly stem from the misfolding of proteins even in the absence of purely genetic causes. This undermines the notion that the sequence of amino acids in a protein uniquely determines its consequent shape and function, and weakens the linkage between genetic information and phenotypic traits exhibited by a cell or organism. Nurture—which causes cellular changes during an organism's life—matters.

Looking forward, computational tools will continue to play a role in pursuing an understanding of both nature and nurture and how they interact.

5. Approaching Protein Folding Sideways

Researchers would love to be able to accurately predict a protein's shape simply by knowing its amino acid sequence. That's because a protein's shape can yield an understanding of the protein's function and reveal the likelihood it will bind to other molecules, including drugs.

Naturally occurring proteins typically fold quickly *in vivo* into the correct, native, functional form. Yet predicting the complete, tertiary structure of proteins by using only sequence information and a reliable force field has challenging chemical, geometric and combinatorial aspects. It may be an intractable NP hard problem that requires approximate solutions.

Rather than attacking the problem head-on, researchers have spent the last decade developing alternative strategies—using known fold motifs, sequence homologies and the many existing structures in the Protein Data Bank (PDB)—to piece together a reasonable guess for the folded form that can be further refined and improved by additional information and calculations.

Though we've come a long way toward efficient handling and quick access to protein structural information via a large variety of different types of queries of massive structure/sequence data, there remains an ongoing need to efficiently and reliably exploit these data in order to predict accurate shapes.

Having determined a protein's accurate three-dimensional structure, computational researchers can move on to the task of determining its function in the cell, including how it cooperates with larger assemblies in biological systems. In the past decade, for example, coarse-grained molecular dynamics simulations have revealed how various proteins interact with the cell membrane. Moving forward, as greater computational power comes on line, we can expect larger and ever more detailed simulations, which will reveal valuable clues to the workings of proteins and cells.

6. Untangling Networks

Over the past decade, computational work has helped highlight the extent to which the programming of life relies on complex biological networks. Such networks include the basic metabolic cycles that have been remarkably conserved throughout evolution, as well as more complex networks that regulate various cell functions. The onset of many diseases, including cancer, is often related to a malfunction of these intra- and inter-cellular commu-

nication networks. Going forward, computational modeling and simulation, together with experimental efforts to decode complex feedback loops (the hallmark of regulation), are likely to play a major role in generating a deeper understanding of biological networks at all levels.

7. Tackling the Brain: From Artificial Intelligence to the Connectome

Computer programs inspired by the central nervous system and designed to “learn” from their environments have been around for some time now. They can steer robots through diverse terrain and are being increasingly successful at tasks such as computer vision and speech recognition. In recent years, some researchers have even attempted to replicate the behavior of neurons and the brain on a computer, with the intention to both gain an understanding of the brain and build more powerful computers.

Advances in computationally intensive imaging modalities, such as functional MRI and diffusion tensor imaging, have also increased our understanding of the brain's anatomy in health and disease. Unlike computers, which store data and perform operations at specific locations within their chips, we now understand that both memory/data and operations in our brain seem to be distributed over many neurons and synapses. It is this richness that may underlie the phenomena of associative memory and thinking. Patterns of closed loop sympathetic multi-neuron firing may eventually prove to be the basis of consciousness.

Among the most exciting developments of the last decade are the emerging efforts to map the human “connectome”—the connections among all of the neurons of the brain. Since the human brain network is more complex than that of the entire worldwide web, its complete mapping will likely be a challenge for the next decade or two. Like the human genome

Since the human brain network is more complex than that of the entire worldwide web, its complete mapping will likely be a challenge for the next decade or two.

project of the 1990s, such a connectome project may mobilize a concerted effort and accelerate its achievement. Though we will likely face many challenges in taking on such a project, the potential benefits to be gained from a deep understanding of the (mis)function of the human brain and mind are extraordinary.

8. An Explosion of Computing Power: More Is Not Enough

Large-scale genome sequencing (and the shotgun method in particular) would be impossible were it not for massive computer power and associated efficient, fast algorithms for sequence alignment. Likewise, molecular dynamics simulations of large biological molecules, recently honored with a Nobel Prize in chemistry, depend on vast computational resources. So too does the dawning age of big data, with its need for efficient storage and quick access in order to probe for patterns and clues and gain a more comprehensive understanding of biological molecules, cells, tissues and organisms.

Thanks to increasing miniaturization, computing power has increased exponentially. However, the variety and complexity of biological information is growing much faster than computational power. Many researchers deal with computational limits by varying the computational resolution of their models and simulations as the scale of the task increases. But the desire to understand biological complexity in all its glorious details makes that approach less than satisfying.

Barring a true breakthrough that provides massive amounts of computational power more efficiently, we will soon be inundated with un-analyzable and therefore largely useless amounts of information. Most potential solutions remain, at this point, more hypothetical than practical—including quantum computers, light-based computers, or an unforeseeable but profound theoretical insight that can vastly improve algorithmic speed. I am somewhat more hopeful that the machinery of DNA/RNA/proteins may prove useful in computing. An all-purpose universal DNA computer likely remains too far in the future, but researchers are already turning to less ambitious lower-level utilizations, such as having complementary DNA probes efficiently search large DNA libraries. When and if any of these possibilities materializes, biomedicine will be ready with the data to take advantage of it.

9. Moving Forward with Molecular Prosthetics: From Synthetic Biology to Nanobiology

Just as computation has been instrumental in designing large-scale prosthetics, such as artificial hip bones, so too is it proving valuable for designing and

synthesizing potential molecular bio-prostheses. For example, synthetic biologists are designing robust substitutes for certain amino acids, which can still be integrated into living proteins. The advent of nanotechnology and nanobiology may further close the enormous gap between large- and small- scale prosthetics by allowing useful intimate interfacing of biological and hardware components at many intermediate scales.

Imagine, for example, the creation of nanocircuits using sticky-end unpaired single stranded portions of DNA that can spontaneously attach with high specificity to complementary ends of other double helical chains; or the induced self-assembly of multi-cellular biological structures that can in turn control the assembly of various other nano-hardware pieces designed to interact with it; or novel materials such as the magical graphene, bucky-balls and carbon nano-tubes being integrated into tissues to collect electronic impulses. It is even conceivable that bionic ears could be designed to increase the range of audible acoustic waves.

Such possibilities raise societal questions as well: For example, to what extent do we want nanomachines roaming our bodies? But it's clear that as these issues are being explored, computation will play a role.

10. Confronting the Complexity of Cancer

No single field of medical research manifests more clearly the diverse, distributed and multifaceted nature of biological information than the study of cancer and cancer therapy, making it an excellent testbed for computational approaches. Indeed,

No single field of medical research manifests more clearly the diverse, distributed and multi-faceted nature of biological information than the study of cancer and cancer therapy, making it an excellent test-bed for computational approaches.

the last decade has seen computational work contribute toward a better understanding of cancer on numerous fronts—from basic biology to diagnosis and treatment.

Computational researchers have worked hand in hand with experimentalists to map the pathways

and biological networks associated with the initiation, growth and spreading of cancer, as well as the critical junctions where these can be blocked via appropriate drugs or drug cocktails. High-end molecular dynamics simulations have started to reveal the mechanisms by which several oncogenic mutations in key cancer-related proteins (such as Ras and p53) wreak havoc in the cell.

On the diagnostic side, scores of biomarkers for many common forms of cancer have been identified with the help of computation and are being widely used. Since cancer is most likely induced by synergistic pathways, further work is needed to determine whether clusters of markers may jointly serve as more reliable indicators.

Insofar as treatment and therapy are concerned, novel combinations of drugs are being suggested for experimental testing based on computational screening. At the same time, physics-based techniques are helping clinicians deliver more accurate and localized radiation to cancerous cells.

And then there is the Big Data organizational and analytical effort to identify patterns of driver mutations and genes, which is beginning to tap into the vast data from cancer patient cohorts in electronic medical records.

Despite this progress, there remains a long road ahead. We need to computationally combine and integrate data types across a range of scales with the aim of predicting which mutational combinations will be oncogenic, and which drugs will benefit individual patients. The successes of the last decade have laid the groundwork for such an effort.

11. Decoding the Microbiome

The last decade saw our initial efforts to decode the microbiome—the microscopic life that resides within our own bodies. The importance of this internal ecological niche to human health is becoming increasingly clear. The microbiome includes not only well-known disease-causing microbes, but also thousands of species of bacteria, such as *E. coli*, with which we have a symbiotic relationship.

The discovery that transplanting specific bacteria

from fat individuals to lean ones and vice versa can transfer tendencies toward obesity or reverse it suggest a much more subtle and deeper layer of interplay between humans and the huge variety of species inside them. Understanding how this phenomenon works at a molecular level will require intensive computation and may aid in our fight against obesity as well as other diseases. It could also provide a diagnostic tool, as enhanced populations of specific bacteria may signal disease onset as well as provide potential clues for treatment.

The time for directly channeling the vast diversity of life inside our bodies to our own medical advantage has come.

12. Building Life from Scratch: From Life's Origins to Synthetic Biology

With the various burgeoning datasets now available, computational researchers are digging into the origins of life as well as designing new forms of synthetic life capable of performing novel tasks.

Researchers are simulating computational models of various evolutionary theories to predict which are most realistic, including theories of how self-replicating molecules arose from a primordial mix of organic molecules; whether RNA or proteins—or for that matter, metabolism—came first; and how proteins have evolved. They are even exploring whether alternative evolutionary schemes are possible on Earth or elsewhere in the cosmos.

At the same time, the last decade has seen the launch of synthetic biology as a powerful field. Achievements include computational models that predict the minimum genome required to support life and validation of that prediction by inserting the genome into a cell to produce a self-replicating organism. In addition, researchers are designing engineered organisms capable of sensing environmental toxins or producing biofuels.

As these fields progress, their revolutionary implications will likely amaze us in ways I cannot begin to imagine. □