In silico models for cellular and molecular immunology: successes, promises and challenges

Arup K Chakraborty, Michael L Dustin & Andrey S Shaw

Although most immunologists do not routinely combine experiments with theoretical and mathematical modeling, new insights can be gained from this interdisciplinary approach. But what makes a good theoretical model?

As with all scientific disciplines, advances in immunology are driven by experimental observations. Observations yield facts that can be integrated into hypotheses and theoretical models that are amenable to further experimental tests. The ultimate product of this iterative process of making observations and theoretical model building is mechanistic understanding. The creation of models has been essential in many important advances in biology; the most famous example is Watson and Crick's model for the structure of DNA[1]. Exploration of ideas with conceptual models before the availability of experimental data has also led to the development of some of the basic paradigms of immunology. Burnet's model[2] for self-nonself discrimination was later confirmed by Medawar's experimental observations[3], and the two shared a Nobel Prize for this advance made using 'theory' and experiment. Similarly, Bretscher and

Arup K. Chakraborty is at the Departments of Chemical Engineering and Chemistry, University of California, Berkeley, California 94720, USA and the Physical Biosciences and Materials Sciences Divisions, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. Michael L. Dustin is at the Skirball Institute of Biomolecular Medicine and Department of Pathology, New York University School of Medicine, New York, New York 10016, USA. Andrey S. Shaw is at the Department of Pathology and Immunology, Washington University School of Medicine, Box 8118, 660 South Euclid, Saint Louis, Missouri 63110, USA.
e-mail: arup@uclink.berkeley.edu

Cohn[4] proposed the existence of a helper T cell, before there was any observational evidence for it, to explain why autoimmunity is so rare. In a similar vein, Lafferty and Cunningham[5] proposed the idea of costimulation to explain certain observations before direct evidence for it was available. Recent years have also witnessed the successful synergistic use of quantitative theoretical models and experiments to elucidate virus dynamics, particularly that of human immunodeficiency virus infections[6].

In the past three decades, advances in cell culture techniques, immunochemistry, recombinant DNA methodology, X-ray crystallography and the use of transgenic mice have enabled the explanation of diverse immunological phenomena in terms of protein structure and specific biochemical events. With a few notable exceptions (such as serial triggering[7] and kinetic proofreading[8]), quantitative models have made a minor contribution to this revolution in cellular and molecular immunology. However, there are some recent indications that quantitative theoretical models can partner fruitfully with cellular experiments to substantially advance our understanding of molecular mechanisms that underlie key events such as T cell activation, T cell migration and thymocyte selection. The use of computational models as essential complements to *in vivo* and *in vitro* experimentation is not a new paradigm. It is a continuation of the traditions established by Burnet, Bretscher, Cohn, Medawar and others, further buttressed by the power of modern computers
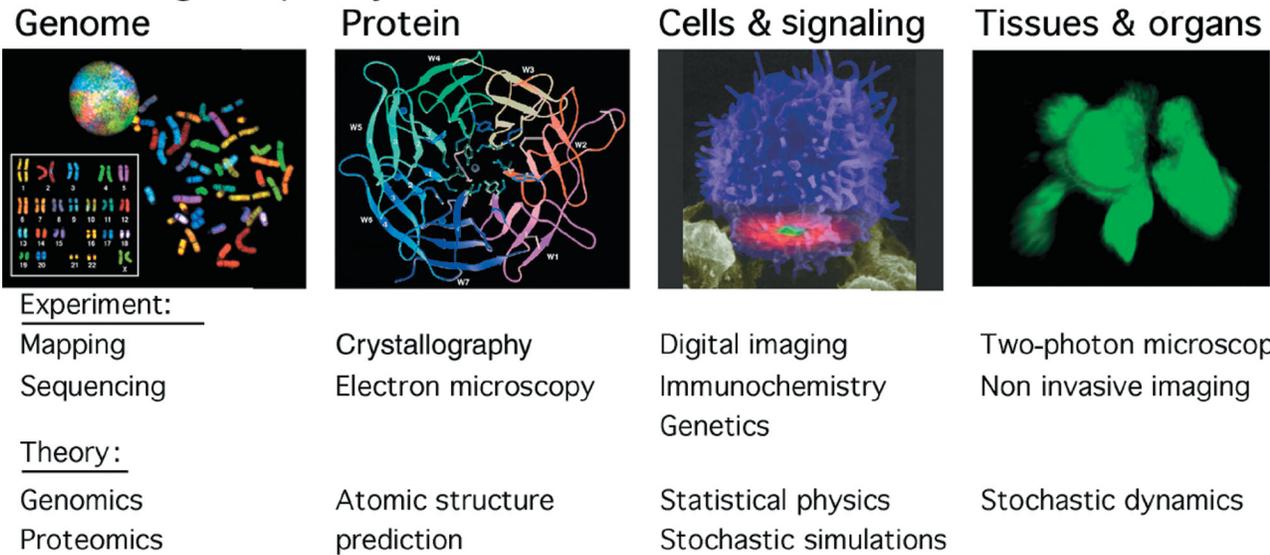
and theoretical methods designed to 'interrogate' complex systems.

Interest in *in silico* models has been catalyzed by many recent imaging studies[9–14]. These experiments allow direct visualization of the large-scale spatiotemporal evolution of cell surface receptors and intracellular signaling molecules during T cell activation and thymocyte selection. These vivid images, combined with the ability to 'count' individual molecules[14] and carry out genetic and biochemical experiments, provide information on complex phenomena that result from cooperativity between many interacting cellular components. It is often difficult to parse the interplay between the many variables involved from the experimental data alone. In particular, the relative magnitudes of competing forces when presented with different stimuli are difficult to assess.

Quantitative models (either analytical or those that can be simulated on a computer) can complement experiments to delineate the effects of different variables. This is because the activity of each element in the system can be manipulated and monitored individually *in silico* with relative ease. The consequences of different mechanistic hypotheses regarding the function of a specific molecular element, or how changes in external stimuli or changes in activity and expression level of a specific molecule influence a cascade of cooperative events, can thus be assessed. In other words, *in silico* models can be used as 'digital analogs' of transgenic animals in which the activity of the immune system can be manipulated in controlled conditions. In fact, computational

Figure 1 Hierarchical models for studying the immune system. Increasing 'complexity' corresponds to situations in which the pertinent phenomena occur on larger length scales and in longer time scales. Bottom, examples of synergistic experimental and theoretical or computational methods that are best suited for studying these phenomena.

models can be manipulated in a more subtle way than transgenic animals can, as any one of many functions of a specific molecule can be eliminated and its consequences can then be studied. This is particularly useful for discriminating between various possible ways a molecule influences a cascade of events. In a computational model, each possibility can be implemented independently, and its effect on cellular response can be examined. These studies can eliminate certain hypotheses from further consideration and focus *in vitro* and *in vivo* genetic experiments on mechanistic issues that are potentially central to the immunologic response. A good quantitative model allows 'virtual' genetic experiments *in silico* to be done more rapidly than *in vitro* or *in vivo* experiments. However, results of *in silico* experiments depend directly on the model being simulated. Hence, care must be taken in formulating the model.

The creation of a quantitative model begins with the formulation of mechanisms through which various individual events (which typically occur on short length and time scales) influence each other. These hypotheses are guided by biological and physicochemical insight. They can be tested by comparison of computational results with known experimental observations. Reproducing phenomenology observed in experiments is merely the first requirement of an *in silico* model, and the calculation of precise numbers is not a particularly important objective. The hallmark of a good theoretical model is that it provides

mechanistic insight that could not have been gleaned easily from experimental observations alone. This insight should then seed ideas for further experimentation that advance the pertinent idea to a higher level of understanding.

Elucidation of complex cellular phenomena in immunology using synergistic *in vivo*, *in vitro* and *in silico* experiments is an avenue of research whereby physical scientists with longstanding interests in cooperative phenomena can contribute to the development of important principles in immunology. Sophisticated theoretical methods, rooted in statistical mechanics and proven to be useful in studying emergent complexity in multicomponent liquids and membranes, combined with a deep understanding of biology, can shed light on the pertinent issues. Bringing together these two ingredients requires collaboration between scientists trained as immunologists and physical scientists. Below, we discuss a few recent examples in which computational models have proven to be important complements to cellular experiments in the quest for understanding the mechanisms that underlie the function of T cells.

The concepts of serial triggering[7] and kinetic proofreading[8] as well as receptor degradation are commonly invoked to discuss T cell responses to different peptide ligands. However, understanding of the interplay between these phenomena remains incomplete. Part of the problem encountered in resolving these issues is that experiments

that manipulate one phenomenon while leaving the others unaltered are often difficult to design. Coombs *et al.*[15] used a computational model and synergistic experiments with T cells to demonstrate that the ideas of serial triggering and kinetic proofreading are consistent with the observed dependence of T cell response on the T cell receptor (TCR)–peptide–major histocompatibility complex binding kinetics only if triggered receptors can stay phosphorylated and 'marked' for degradation for a certain time after unbinding from the ligand. In their computational model, Coombs *et al.* were able to assess independently the effect of various hypotheses (for example, only bound TCR can be degraded, only unbound TCR can be degraded and so on) on cellular response, and eliminate those that were inconsistent with experimental data. Such highly controlled *in vitro* or *in vivo* experiments are difficult.

The immunological synapse is a stable cell-cell junction that in the case of mature T cell–antigen-presenting cell interfaces is characterized by a central accumulation of TCRs (called the central supramolecular activation cluster, or C-SMAC) surrounded by a peripheral ring of adhesion molecules (called the P-SMAC)[9,10]. Although both experiments and theoretical models have shed light on the forces that drive synapse formation during T cell activation and thymocyte selection[9–11,13,14,16], the biological function of the synapse has been controversial. Early studies

postulated that the C-SMAC enhances and sustains TCR-mediated signaling for the long times necessary for commitment to activation and proliferation[9,10]. More recently, the paucity of active signaling molecules in the C-SMAC after a few minutes led to the postulation that the C-SMAC is not involved in signaling[12]. The many coupled events involved in signal transduction and receptor clustering made it difficult to refute or assert either point of view. Recently, computational studies and genetic experiments[17] were delicately intertwined to delineate the various effects and resolve the controversy surrounding TCR signaling in the C-SMAC.

Lee *et al.*[17] formulated an *in silico* model for intracellular signaling and C-SMAC formation in T cells. The model is sufficiently complicated in that it includes the essential aspects of TCR based signaling. At the same time, it is not so complicated that the results of simulating the model reflect specific choices of parameters or features of the network that are not well established facts. Careful monitoring of the dynamic events during *in silico* experiments in which the C-SMAC formed and those in which C-SMAC formation was disabled showed that the C-SMAC is a site of enhanced receptor triggering. Why then is there a paucity of active signaling molecules in the C-SMAC at longer times? Computer simulations showed that because only triggered TCRs are subject to degradation, enhanced production of triggered TCRs in the C-SMAC also results in a higher rate of receptor degradation. An experimental system was then designed to test the prediction that the C-SMAC is a site of intense TCR triggering.

CD2AP-deficient T cells show defective receptor degradation and form C-SMACs with supported lipid bilayers containing peptide–major histocompatibility complexes[17]. The computational model predicts that at later time points, the strongest signaling should be found in the C-SMACs formed by CD2AP-deficient cells, as receptor clustering in the C-SMAC enhances TCR triggering, but concomitant receptor degradation cannot occur. In contrast, wild-type cells showed the weakest signaling in the C-SMAC at similar time points[12]. Consistent with this prediction, *in vitro* experiments show that after 60 minutes, the strongest phosphotyrosine activity is found in the C-SMAC of CD2AP-deficient T cells, but wild-type cells have the strongest activity in the surrounding P-SMAC[17]. These investigations show that the C-SMAC enhances receptor triggering, which leads to a concomitant enhancement in receptor degradation. Thus, the C-SMAC

seems to be a motif that boosts TCR signaling but is also self-limiting to prevent antigen-induced cell death. This hypothesis that the C-SMAC is an adaptive controller of the strength and duration of signaling now needs to be tested further (for example, with experiments with altered peptide ligands).

The synergy of molecular genetic and *in silico* analysis is also demonstrated by the study of Hoffman *et al.* on the activation of NF-κB[18]. NF-κB represents a family of transcription factors that is essential in inflammatory processes and immune cell activation through coordinated transcriptional regulation of genes encoding chemokines, cytokines and other molecules. NF-κB is held in the basal state in the cytoplasm by three competing inhibitory factors: IκBα, IκBβ and IκBε. Signal transduction from specific surface receptors activates IκB kinases, which phosphorylate IκB isoforms, leading to ubiquitination-dependent degradation and nuclear translocation of the NF-κB transcription factors. The gene encoding IκBα is itself transcriptionally up-regulated by NF-κB, leading to feedback inhibition, whereas the production of IκBβ and IκBε is constitutive. The way in which these three inhibitors contribute to effective biological regulation was not clear. A computational model predicted that feedback inhibition by IκBα could lead to rapid down-regulation of NF-κB after transient stimuli, but that oscillation of nuclear NF-κB could occur with sustained stimuli. The modeling further predicted that the constitutively expressed IκBβ and IκBε cannot rapidly down-regulate nuclear NF-κB after transient stimuli, but should dampen oscillations during sustained stimulation. Considerable agreement of these predictions with experimental results was obtained for the response to tumor necrosis factor of all combinations of double-IκB-deletion fibroblasts. The results of this computer-guided experimentation may also help explain aspects of regulation of the gene encoding interleukin 2 in T cells in response to sustained versus transient cell-cell interactions.

Another example of synergy between modeling and experiment does not involve immune cells or even animal cells, but has implications for lymphocyte migration nonetheless. Spontaneous cell polarization in the absence of external or pre-existing internal queues is an important process in many types of cells of higher animals, but is not readily accessible to genetic analysis. Understanding spontaneous polarization is important for immunology, as it seems to be the driving force behind immune surveillance in lymph nodes[19]. In addition, the evolutionarily

conserved G protein Cdc42 is specifically linked to spontaneous polarization of T cells[20]. Wedlich-Soldner *et al.*[21] established a genetic model for spontaneous polarization of the actin cytoskeleton in the yeast *Saccharomyces cerevisiae*, and then formulated and tested a mathematical model for the establishment of cell polarity based on a positive feedback involving actin filament–directed exocytosis of vesicles coated with active Cdc42. The essential contribution of the model in this system was to generate a quantitative and testable prediction of a specific molecular hypothesis, which could then be tested with tools of genetics and cell biology.

The examples described here indicate that computational models can be useful complements to genetic, biochemical and imaging experiments. These successes should encourage further synergistic experimental and computational studies aimed toward elucidation of the complex biology of T cell activation and thymocyte selection. Indeed, it is possible that future progress could involve the use of accurate computational models ('digital transgenic mice') for simulation of the adaptive immune response. Such computational models would have to be hierarchical, as they need to encompass phenomena that occur over a wide spectrum of length and time scales (**Fig. 1**). 'Atomistically detailed' models that can study events that occur on the scale of less than a nanometer and in time scales faster than picoseconds are synergistic with X-ray crystallography. Information from atomic-scale models will serve as input to field theoretic models and associated computations that study the cell as a system of interacting components. These studies, which yield results on phenomena occurring on length scales of micrometers and time scales of seconds, minutes and hours, are synergistic with genetic, biochemical and imaging experiments. Results from cellular-scale models subsequently become inputs to larger-scale models that can study cell migration and trafficking in synergy with experiments such as real-time images of thymocytes and T cells in tissues[19,22,23]. Finally, such models provide inputs to models for virus dynamics, the spread of infections and associated epidemiological studies[6].

In conjunction with experiments, such comprehensive hierarchical models could potentially be used to understand the cause of pathologies, suggest drug targets and simulate the effects of drugs designed to correct pathologies. They may also contribute to our understanding of the molecular and cellular processes that lead to effective self tolerance and help develop strategies that can break self

tolerance to destroy tumors and chronic infections. Recent molecular advances, including the identification of 'master regulatory genes', indicate that the molecular building blocks of the immune system will soon be known. Understanding how these units fit together and interact is a problem of herculean scale that may be best 'hefted' by combined computational and experimental approaches.

The creation of computational models that can aid the development of modalities for controlling the immune response is faced with substantial challenges, however. The formulation of hierarchically arranged models that seamlessly pass information from one level to another has not been accomplished and faces considerable hurdles. Stoichiometries, expression levels and detailed kinetic information on reactions involved in signal transduction are not known, and accurate measurements of these quantities are needed. The structure and function of many important multiprotein signaling complexes remain undetermined. Confronting and overcoming each of these challenges would be easier if experimental and theoretical studies were done in synergy. This will require that computational scientists trained in the physical sciences and immunologists make an effort to lower communication barriers between them. Rich dividends could accrue from bringing together two major advances of the latter half of the twentieth century: molecular and cell biology and computer simulation of complex systems.

1. Watson, J. & Crick, F. *Nature* **171**, 737 (1953).
2. Burnet, F.M. *The Clonal Selection Theory of Acquired Immunity* (Vanderbilt Univ. Press, Nashville, 1959).
3. Billingham, R.E., Brent, L. & Medawar, P.B. *Nature* **172**, 603–606 (1953).
4. Bretscher, P. & Cohn, M. *Science* **169**, 1042 (1970).
5. Afferty, K.J. & Cunningham, A. *Aust. J. Exp. Biol. Med. Sci.* **53**, 27–42 (1975).
6. Perelson, A.S. *Nat. Rev. Immunol.* **2**, 28–36 (2002).
7. Vallitutti, S., Muller, S., Cella, M., Padovan, E. & Lanzavecchia, A. *Nature* **375**, 148–151 (1995).
8. McKeithan, K. *Proc. Natl. Acad. Sci. USA* **92**, 5042–5046 (1995).
9. Monks, C.R., Freiberg, B.A., Kupfer, H., Sciaky, N. & Kupfer, A. *Nature* **395**, 82–86 (1998).
10. Grakoui, A. *et al. Science* **285**, 221–227 (1999).
11. Ritchie, L.I. *et al. Immunity* **16**, 595–606 (2002).
12. Lee, K.H. *et al. Science* **295**, 1539–1542 (2002).
13. Hailman, E., Burack, W.R., Shaw, A.S., Dustin, M.L. & Allen, P.M. *Immunity* **16**, 839–848 (2002).
14. Irvine, D.J., Purbhoo, M.A., Krogsgaard, M. & Davis, M.M. *Nature* **419**, 845–849 (2002).
15. Coombs, D., Kalergis, A.M., Nathenson, S.G., Wofsy, C. & Goldstein, B. *Nat. Immunol.* **3**, 926–931 (2002).
16. Qi, S.Y., Groves, J.T. & Chakraborty, A.K. *Proc. Natl. Acad. Sci. USA* **98**, 6548–6553 (2001).
17. Lee, K. *et al. Science*, published online 25 September 2003 (doi:10.1126/Science.1086507).
18. Hoffmann, A., Levchenko, A., Scott, M.L. & Baltimore, D. *Science* **298**, 1241–1245 (2002).
19. Miller, M.J., Wei, S.H., Cahalan, M.D. & Parker, I. *Proc. Natl. Acad. Sci. USA* **100**, 2604–2609 (2003).
20. del Pozo, M.A. *et al. Eur. J. Immunol.* **29**, 3609–3620 (1999).
21. Wedlich-Sodner, A. *et al. Science* **299**, 1231–1235 (2003).
22. Stoll, S., Delon, J., Brotz, T.M. & Germain, R.N. *Science* **296**, 1873–1876 (2002).
23. Bousso, P., Bhakta, N.R., Lewis, R.S. & Robey, E. *Science* **296**, 1876–1880 (2002).