Sense of Self
A Synopsis of the Immunotolerance Debate

by Moderator Kenneth F. Schaffner

(Issue 11 · posted June 27, 1997)

The nineties have witnessed a vigorous and still-emerging debate about the fundamental nature of immunology. Though much of this has occurred in the professional journals and out of public view, readers of the science section of the *New York Times* were startled by a headline in March 1996 stating "Findings Pose a Challenge to Immunology's Central Tenet" (Johnson, 1996). That "central tenet" is the view that recognition and preservation of "self" from "nonself" is what the immune system was selected by evolution to do; the *Times* was reporting on some just-published articles in *Science* arguing for a contrary "danger" recognition view. The implications of a possible overturning of a central concept of immunology are potentially vast: from new ways to research vaccines and control infectious diseases (including HIV/AIDS), to novel understandings of cancer and autoimmune diseases and new therapies, to better ways to keep transplanted organs from being rejected by the recipient. Serious doubts have been expressed, however, about this turn of events (Silverstein, 1996; Miller in Pennisi, 1996).

Over the course of the past year, an increasingly vocal (and visible) debate has developed, with further research done on the danger theory, championed by Fuchs and Matzinger, and on other major alternatives as well. These contrasting approaches include extensions of Jerneian idiotype network theories by Coutinho, Stewart, and Bandeira, as well as "stranger," morphostasis, "integrity," cytokine-cascade, and antigen-localization accounts by Janeway, Cunliffe, Dembic, Weigle, and Zinkernagel, respectively. The associative-recognition model of Bretscher, Cohn, and Langman has also evolved in this context.

These events set the stage for the Cutting Edge debate held in May, in which many of the major proponents of competing theories could actively present their views and criticisms of alternative approaches. The participants in this stellar group each suggested several possible questions and themes for debate and then slugged it out over the course of a week. As a historian and philosopher of immunology, but not a proponent of any specific theory, I was asked to serve as the reasonably objective moderator. At the end of each day, I attempted to summarize where the discussion had gone that day and posed what seemed to be the next logical question to the participants.

The participants included one of the originators of the "danger" theory, Ephraim Fuchs from Johns Hopkins; the proponent of the "integrity" theory, Zlatko Dembic of the University of Oslo; and a spokesperson for the "network" theory, Antonio Bandeira at the Institut Pasteur. Also involved in the debate were the Scripps Institute's Bill Weigle, an experimental immunologist and proponent of an experimentally based cell-cytokine approach; Doug Green from the La Jolla Institute for Allergy and Immunology, who initially tried to take a nontheoretical stance but ended up arguing against a self-nonself distinction and in favor of a "local damage" theory; and Rod Langman from the Salk Institute, a vigorous defender and developer of the Bretscher-Cohn-Langman associative-recognition (and self-nonself) model. Melvin Cohn of the Salk Institute did not participate but did observe the debate, and he provided some valuable background to the moderator prior to the debate. Both Polly Matzinger from the NIH and Arthur Silverstein from Johns Hopkins were asked to participate, but could not because of travel commitments. Arthur Silverstein did circulate a draft essay by him and his colleague Noel Rose on the debate topic to the participants.

The debate began with a general question asking what were the main functions of the immune system; or, to put it another way, why was the immune system developed by evolution? This question was designed to look for whatever consensus we felt might emerge before polar positions were taken on the viability of the self-nonself distinction and on the competing models. These two issues occupied us the following two days, with the most extensive series of postings (47!) occurring on Day 3, which dealt with the specifics of the

competing models. Day 4's question asked the participants to look more closely at novel experimental evidence, classical experiments from their perspective, and potential clinical implications of the new and developing views. Day 5 asked how we might choose between the theories and whether yet-to-be-done experimental results could help to make that choice.

On Day 1, the answers were fairly general, with details and mechanisms (both cellular and molecular) left until later. Doug Green's answer, that the immune system developed "to maximally damage parasites while minimally damaging us" (#1), was agreed upon to a large extent; participants also urged an "integrity preservation" addition (#2) and consideration of how tolerance is maintained in adult life (#3). The need for a parasite damager (and "ridder") was introduced, as well as the need to distinguish self from nonself (#4). Ephraim Fuchs noted that the immune system evolved to react to "non-apoptotic death or cell stress" and also to lots of apoptosis (a "gentle" and apparently "altruistic" cell suicide of sorts) (#5). Antonio Bandeira presented an argument based on a hypothesis of the evolution of variable-region molecules to support the self-nonself preservation view (#9; further elaborated in messages 11 and 13).

Hints of a dispute over the self-nonself distinction were brought into the open on Day 2. First, it was suggested that self-nonself was a good approximation of the truth but that the danger and integrity approaches were better (#1 and #3). It was argued, however, that if the immune system can kill pathogens and not the host, it *must* make a self-nonself distinction, and maybe the language of self-nonself should be replaced with something less freighted, though the issue would be the same (#2). The self-nonself distinction was noted to have been useful in immunology as long as it generated novel and testable theories of tolerance, which it had done in the hands of Burnet, Lederberg, and Bretscher and Cohn (#5). Bill Weigle stated that self-nonself was useful as long as we recognized its limitations but that we should not be looking just for *one* general mechanism (#8). Doug weighed in with what he said would be an "inflammatory" view: that self-nonself "is no longer useful to the study of immunology" and that we would do better to look at a more specific level (#11). According to Antonio, if we understand it right - that is, from the network point of view - the terms "self" and "nonself" are acceptable (#25).

To sharpen the discussion, Rod Langman proposed a thought experiment involving Greenian "thingies" (#9) (like the simple ones proposed by Doug Green in Day 1, message 1), which led to an elaboration of the nature of signals in the integrity theory (#26), along with a lot of "alphabet soup." One recurring theme in this and other days was the question of whether the "self" continually changes throughout life. At the end of the Day 2 discussion, a consensus of sorts was detected, as well as some progress (#32).

The real meat of the debate can be found in Day 3, although Days 4 and 5 went on to elaborate on and apply in important ways the substantive positions developed in Day 3. Professional immunologists will find the concise statements of the alternative positions, and the sharp criticisms of competing views, of particular interest.

Rod began the Day 3 discussion with "a big sitting duck" called the "webbed associative antigen recognition" model - wAAR, or "war" for short (#1). He also covered historical ground, including the Lederberg model and its extensions within the associative antigen recognition (AAR) model. Key to the model are the two pathways of lymphocyte differentiation to a T (thymus-derived) helper call from an indeterminate state (iTh) to an effector state (eTh) and the way that this is used to make the self-nonself distinction operational. Doug sketched an alternative view in terms of cellular and molecular mechanisms that relied on damage and local inflammation to trigger dendritic cells to change (#2). These, when they encounter appropriate T cells, stimulate the T cells to proliferate and react against the source of the original damage. Modified circumstances could also lead to tolerance. Doug claimed that this was not "a theory, or a model, just an amalgam of what a lot of us already know," but to the moderator, it looked a lot like the danger theory.

Next, the integrity protection/restoration theory (#3 and #4) and the danger model (#5) were concisely stated, and the AAR model was criticized. Rod responded to this criticism and assessed what was common among all approaches and what was distinctive about the Coutinho-Stewart-Bandeira, Bretscher-Cohn-

Langman, and what he termed the "alarmist" approaches (which also included the danger model and presumably Doug's and also Janeway's view) (#6).

Bill developed an account that focused on what some key antigens (monomeric human gamma globulin) can tell us about tolerance and on the roles of antigen-processing cells, the cytokine cascade, and upregulation of costimulatory elements in the immune response (#7, #13, #27). The AAR, danger, and integrity models were further argued; Zlatko Dembic claimed that he could incorporate the danger hypothesis within the integrity model (#9), and Rod posed a challenge relating to the breaking of tolerance and suggested that the sequence of responses cannot be explained by either the network or the alarmist theories (including danger and integrity) (#16 and #17). Interaction on this issue led to the conclusion that alarmist theories can explain such a sequence, but the evidence cannot be easily obtained - although the participants now understand what must be measured!

Bill developed an account that focused on what some key antigens (monomeric human gamma globulin) can tell us about tolerance and on the roles of antigen-processing cells, the cytokine cascade, and upregulation of costimulatory elements in the immune response (#7, #13, #27). The AAR, danger, and integrity models were further argued; Zlatko Dembic claimed that he could incorporate the danger hypothesis within the integrity model (#9), and Rod posed a challenge relating to the breaking of tolerance and suggested that the sequence of responses cannot be explained by either the network or the alarmist theories (including danger and integrity) (#16 and #17). Interaction on this issue led to the conclusion that alarmist theories can explain such a sequence, but the evidence cannot be easily obtained - although the participants now understand what must be measured!

Doug (#39) raised some questions for Ephraim regarding the effects of virus on apoptosis, and Zlatko criticized the AAR model by suggesting that it cannot satisfactorily explain unresponsiveness to newly emerging self antigens like the ones at puberty, in antibody idiotypes, or influenced by exchanged bacterial flora (#43). Rod answered that there is no definite proof for any novel self antigen that emerged during the life of an individual and that the AAR model will hold until such proof is obtained (#44); he included a side note (#41) clarifying a confusion in differentiation states that he had earlier introduced. Bill closed this (long) day with an objection to the AAR (#47).

In Day 4, each of the participants offered suggestions, some quite specific, for applications to cancer treatment, autoimmune diseases, and organ transplantation. Readers who work in these more clinical areas may well want to follow up with the participants on moving these vital fields forward into animal models and, ultimately, human testing.

Zlatko led off by addressing experimental and clinical domains in terms of both peripheral and central tolerance (#1). He suggested that cancer cells may escape immune destruction because they mimic integrity signals or downregulate signal[3] in his model. He added that necrotically destroying cancer cells might prime dendritic cells to mount an immune response against the tumor. Zlatko interpreted autoimmune disease as causing an (tissue-specific?) aberrant signal[2] that might be susceptible to costimulation blockers, which also might be useful in preventing organ transplant rejection. Ephraim, who is also a clinical oncologist, identified the two features of the danger model that are relevant to cancer and drew two implications for cancer. He then proposed specific ways to try to generate an immune response against tumors. Finally, he proposed a novel experiment and explained some classical findings using the danger theory, including high and low zone tolerance.

Bill distinguished central (or thymic) and peripheral tolerance within the context of the model he developed on Day 3 and examined ways that autoimmune disease might arise in such a view (#3). He objected to Ephraim's distinction between low zone and high zone tolerance, characterizing it as "dogma" with very slim experimental support (#5). Antonio disagreed with Bill's central/peripheral distinction but emphasized a "huge difference" between perinatal and adult immune systems from the Coutinho-Stewart-Bandeira perspective (#7). He argued further, from this perspective, that autoimmune diseases are *not* a hyperresponse of the immune system but rather should be treated as immunodeficiency diseases, and immunostimulation therapies should thus be considered. He also speculated on how one might address both tumors and organ

graft rejection. Zlatko offered experimental support for Bill's central/peripheral distinction from his laboratory as well as looking at some implications for the integrity theory of those results (#8).

The AAR model was suggested to have some interesting clinical applications to reduce the autoimmune response and to establish tolerance to organ grafts (Langman, #4). Some type of class switching (from cell to antibody mediated, or among antibody types) might be explored to induce unresponsiveness. One might also look for ways to block the expression of eTh function and permit them to revert to iTh state. Noting that his views on Day 3 may have looked a lot like a danger model, Doug offered some differences focusing on local responses. He added that, in autoimmunity, we probably focus too much on lymphocytes and not enough on specific tissue effects - a kind of "self-marker" idea. He discussed the immune response in the eye and suggested that the eye's main defense mechanism, Fas ligand, which triggers an apoptotic response in virally infected cells, may have broader implications - and perhaps even explain the eventual acceptance of liver transplants. (For some related views not addressed in this debate, see Starzl et al., 1996.)

The responses to Day 5's question ranged from the view that we can test only subsidiary, more specific hypotheses and not general theories in this area, to the Popper-like proposal of two findings that would falsify the danger theory if they were experimentally confirmed (Popper, 1959). It was suggested that because it's hard to prove or disprove some of the proposed models, looking at the details of the nature of the dendritic cell or the mechanism of costimulation might allow some differences between the danger and integrity models to be specified and looked for experimentally (#1). Zlatko reminded us that the answers to questions will vary in their simplicity, suggested that we go from the simplest onward, and provided an overview of how this might be viewed in moving from self-nonself through the Bretscher-Cohn-Langman two-signal model, to a Coutinho-Bandeira network account, then to danger and dangerlike models and his favored integrity model. Bill noted that he has spent considerable time debating theories over the years but that there is no agreement among holders of various theories about experimental proof or disproof (#2). What he did, therefore, was to provide a concise summary of his view of how tolerance is induced and maintained at the cellular and subcellular levels, along with citing specific experimental supporting data. Rod defended the value of models because they help to distinguish principles from details and reinterpreted Bill's account in terms of a second-signal (AAR) model (#3). To distinguish AAR from danger (or more generally "alarm" models), Rod proposed that we need to determine whether AAR is required for the origin of eTh or whether a local alarm is sufficient. For this he returned to the tolerance-breaking discussion of Day 3 (#17, #26, #29, and #32-38) and also proposed what might be used to distinguish AAR from the dominant-suppression network model.

Ephraim offered two experiments that would *disprove* the danger theory, because he believes it is never possible to prove a theory (#5). These depend on the two "key" features of the danger model: (1) only "professional" antigen-presenting cells (APCs) can initiate an immune response in naive T cells and (2) APCs must be activated by an endogenous or exogenous danger signal. Doug argued that we can reasonably test hypotheses (e.g., that interleukin-2 is required for T-cell proliferation) in immunology but that theories are harder - more "slippery" - to test because a theory can be modified without really giving it up (#8). (For discussion of the modifications the clonal-selection theory went through on just this point, see Schaffner, 1993) Theories were noted to be valuable because they generate interesting and testable hypotheses; rather than try to test one theory against another, Doug recommended that we study a theory to see what new questions it can generate and then test them. Rod disagreed, arguing that we need to test "well-formulated" theories lest we, and the rest of immunology, produce vast amounts of data that make little difference to real progress (#10).

Overall, none of the debaters convinced any of the others to give up his favored theory, but the differences between the theories, as well as some common families of theory such as the "alarmist" class, were identified. Day 5's discussion indicated in fairly explicit ways why changing minds is, and should be, difficult. A consensus emerged over the course of Days 1 and 2 that the debate had to move to the level of specific models (as it did in Day 3), but disagreement continued as to whether higher-level principles, such as self-nonself or integrity, added to immunologic knowledge at this point. A particularly valuable aspect of the interchange was the clear and concise formulation of the various competing models in Day 3 and the follow-up in Days 4 and 5. Several new experiments or experimental directions were proposed for further

study during Days 2-5, and Day 4 gave these prominent theorists in immunology a chance to articulate some potentially momentous clinical implications of their various models.

Endlinks

A Brief Guide to Peripheral Tolerance in the Immune System - readers with little background in immunology will find this a helpful debate accompaniment. The site includes a glossary.

ClinWeb - extensive links to Web sites dealing with immunologic diseases and the anatomy of the immune system.

The *Journal of Immunology*'s Cutting Edge is accessible for free. Rapid Communications of novel research and current issues in immunology, contains mechanism for dialogue with authors.

The Journal of Immunology - available online. Readers may post feedback to, and receive feedback from, authors.

Turned on by Danger - a BBC film by Michael Mosley on Dr. Polly Matzinger and her danger model, broadcast April 17, 1997. The site includes a full transcript, images from the film, and commentary from various experts in the field of immune tolerance.

PNAS Online - has a sampling of recent commentaries dealing with immunological tolerance, including:

- Fas-ligand: Privilege and peril by Douglas R. Green and Carl F. Ware
- The mother-child union: The case of missing-self and protection of the fetus by Wayne M. Yokoyama