REVIEW

The Art of the Probable: System Control in the Adaptive Immune System

Ronald N. Germain*

The immune system provides very effective host defense against infectious agents. Although many details are known about the cells and molecules involved, a broader "systems engineering" view of this complex system is just beginning to emerge. Here the argument is put forward that stochastic events, potent amplification mechanisms, feedback controls, and heterogeneity arising from spatially dispersed cell interactions give rise to many of the gross properties of the immune system. A better appreciation of these underlying features will not only add to our basic understanding of how immunity develops or goes awry, but also illuminate new directions for manipulating the system in prophylactic and therapeutic settings.

The immune system provides the organism with the ability to recognize, respond to, and in most cases successfully defend against a wide variety of infectious agents. Because they are in a race with the replicative capacity of microorganisms, immune responses must not be "too cold," meaning too insensitive in initiating, too slow in developing, or too meager in expanding, so that effector function should not be overwhelmed by rapidly replicating pathogens (*1*). On the other hand, they cannot afford to be "too hot," because the mediators meant to combat the invader are themselves capable of substantial destructive effects on host tissues. The responses must be "just right"—rapid, vigorous, properly modulated, and of the correct quality (*2*). Operation of a relatively strict set of rules would seem necessary to ensure well-controlled behavior by such a complex system, resulting in a very mechanistic interpretation of most immunological experiments—input X directly and reliably causes a cell to generate output A under a given initial condition.

But as one dissects the immune system at finer and finer levels of resolution, there is actually a decreasing predictability in the behavior of any particular unit of function (a gene, a cell) (*3*–*6*). A major challenge is thus understanding what endows the overall ensemble with both sensitivity and global reliability despite variations in the concentration, initial state, local behavior, or precise number of the participating components. Emerging evidence suggests that these properties arise from a combination of (i) small alterations in the probabilistic behavior of a biochemical pathway, a cell, or a collection of cells (*5*, *7*); (ii) the amplification of these induced changes by positive-feedback loops (*6*) and/or exponential cell growth (*8*); (iii) the action of counter-regulatory controls that modulate these potentially unidirectional, explosive processes; and (iv) the summation of individual topographically dispersed cellular responses subject to these local control mechanisms. Other reviews in this issue will deal with specific molecules and pathways involved in lymphocyte behavior, dendritic cell (DC) function, or vaccine efficacy. Here I will focus primarily on elaborating general principles involved in effective operation of the immune system.

Lessons from Population Biology

Clonal expansion of antigen-specific lymphocytes is a central feature of adaptive immune responses (*9*). In cell populations undergoing such exponential growth, small effects on the behavior of individual cells can be markedly amplified, producing large differences in the final state of the system. This is analogous to the dramatic effects on population frequency seen after several generations when a particular reproductive group has a modest selective advantage (*10*). Unfortunately, the implications of such behavior are often overlooked in drawing conclusions from immunological data. Consider the classic "two-signal" model of T cell activation (*11*), in which CD28 costimulation complements T cell antigen receptor (TCR) input to promote interleukin-2 (IL-2)–dependent proliferation (*12*). Assayed after 5 to 7 days, stimulation by the TCR alone typically results in cell yields that are only a few percent of those seen with both signals. This is often taken to mean that each cell requires the conjoint intracellular signals of these two receptors for activation to a state suitable for cell division. However, suppose that compared to TCR stimulation alone, CD28 cosignaling merely increases the amount of IL-2 made by some cells (*13*). IL-2 concentration has been shown to affect T cell doubling time (*14*). For example, if the greater concentration of IL-2 reached in the presence of CD28 results in a change from a 10-hour cycle to an 8-hour cycle, over 7 days this will produce a 16-fold increase in cell yield. Thus, the difference in output typically seen upon adding CD28 signaling to TCR stimulation can be readily explained by a modest change in cycle time, iterated over many cell divisions [Web fig. 1 (*15*)]. No individual T cell requires two signals to divide or to produce IL-2 at levels necessary for cell division; rather, the probability that many cells will divide more often is increased by costimulation as a result of a minor increase in IL-2 production. The large impact of costimulation on the measured response is dependent on the amplifying effects of clonal expansion rather than a direct effect that licenses the cells for proliferation per se.

Similar considerations apply to experiments dealing with receptor-mediated signaling. Frequently, little importance is attributed to small (<2-fold) differences in the activity of a proximal signaling molecule studied under two different test conditions. However, high levels of nonlinear amplification in a biochemical cascade, akin to the exponential growth of cells in the experiment just described, may convert these minor quantitative differences in upstream events into large downstream changes affecting gene expression. In both this case and that of cell proliferation, a lack of appreciation of these dramatic effects of amplification can easily lead to a failure to consider alternative interpretations for a biological phenomenon [see below and (*16*)].

Autoimmunity Revisited

The potential physiological relevance of thinking in terms of probabilistic behavior, small differences in early responses, and effects of exponential amplification come to light when considering autoimmune responses in genetically prone individuals. Despite the presence of all the necessary susceptibility genes, such individuals vary in whether or not they will develop full-blown autoimmunity (*17*). Both environmental effects and distinct lymphocyte receptor repertoires in otherwise identical individuals may contribute to this incomplete penetrance, but there is another possible source of this heterogeneity that is largely overlooked. Perhaps initiation of disease is a quantal event that depends on the behavior of only a single lymphocyte.

Lymphocyte Biology Section, Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892–1892, USA.

*To whom correspondence should be addressed at the Lymphocyte Biology Section, Laboratory of Immunology, Building 10, Room 11N311, 10 Center Drive MSC-1892, National Institutes of Health, Bethesda, MD 20892–1892, USA. E mail: rgermain@nih.gov

Among mature peripheral lymphocytes in normal individuals a small number of T cells specific for a particular autoantigen presumably exist whose TCR affinity approaches that required for strong signaling. The intracellular signaling competence of these latter cells at any moment will show a Gaussian distribution. Therefore, during any time interval only a small number of all potentially autoreactive lymphocytes will produce an optimal intracellular response even if they are exposed to maximal physiological levels of both processed autoantigen and costimulation. Typically, even the most sensitive of cells achieves insufficient stimulatory signals to overcome the ambient inhibitory effects of regulatory lymphocytes and cytokines (*18*) ("activation barrier" in Fig. 1).

Genetic variation can affect the mean and range of signaling competence, the amount of CD28 ligand available, the number and activity of inhibitory T cells, and/or the ambient levels of suppressive cytokines. Susceptibility alleles presumably move one or more of these factors closer on average to a level suitable for cellular activation, so that occasionally this produces a situation in which the stimulatory signals exceed the inhibitory controls. The result could be that only a single cell "jumps" the activation barrier, begins to divide, and gives rise to progeny that acquire effector competence (Fig. 1). The exponential nature of cell division ensures that only a few cycles are needed to generate a large pool of effector cells. In addition, the TCR signaling apparatus of activated T cells seems to have increased sensitivity (*19*), and effector gene loci are modified so that they can be transcribed with less TCR and cosignaling input (*20*). The consequence is that now many more cells than the single cell whose signaling first exceeded the threshold can continue to respond to the ambient level of self-antigen presentation. The tissue damage caused by these activated T cells can raise the level of presentation of additional self-determinants and promote presenting-cell activation ("danger") (*21*), producing conditions that push other naïve cells into the response (*22*).

This scenario emphasizes that the difference between health and disease could be the "stochastic" activation of a single cell, followed by positive feedback in the form of a gain in TCR sensitivity and multiplication of the responding cells to high numbers. Genetic risk factors for autoimmunity may operate by lowering the barrier the cell must jump to start the cycle, allowing fluctuations in signaling, regulation, and/or antigen presentation to yield occasional breaches that initiate the autoimmune process. The efficacy of brakes on the expansion and function of the initially activated cell pool [e.g., CTLA-4 (*23*) or PD-1 (*24*) inhibitory pathways for activated T cells] may also be diminished by susceptibility genes so that these checkpoints can be bypassed by one of the accumulated activated cells. Thus, the quantal event could be the very early one described above, or the transition of an activated cell within the expanded population from a tissue-infiltrating to a tissue-damaging state (*25*). This "rogue cell" model of the initiation of autoimmunity has substantial implications with respect to the possible use of prophylactic immunosuppressive strategies in genetically susceptible individuals [Supplemental note 1 (*15*)].

System Properties and Circuit Components

Of course, under normal conditions, the type of stochastic disruption of immune homeostasis just outlined does not typically result in disease. How then does the immune system carry out its role in host defense in a reliable and effective manner while avoiding such pathological responses? Although sensitivity is clearly one key attribute, the system must also show a high degree of selectivity, reacting to appropriate signals yet being unresponsive to irrelevant stimuli. Having made a choice, it should be capable of sustaining this decision and suppressing a switch to other possible states. And as a whole, it should maintain a balance that prevents the overall system from locking irretrievably into a single mode when multiple distinct responses might be needed for physiological homeostasis. Finally, it must do all these things reproducibly when the amounts and activities of the component elements show a wide range of absolute values ["robustness" in a system (*26*)].

Feedback pathways promote and regulate sensitivity. Positive feedback plays a central role in enhancing sensitivity and in driving a response fully toward a polarized state. In systems with this property, accumulation of the product (molecule or differentiated cell) promotes self-generation in increasing amounts, ensuring high levels of the product with only modest initial input [nonlinear amplification (*6*)]. Shutting off this type of pathway requires an extremely strong negative influence on the generation or stability of the end product such that the concentration of the inducer is forced so low that reamplification cannot occur.

Although this type of feedback provides a way of potently amplifying weak inputs, it poses the threat of full activation in response to low-level nonspecific inputs ("noise") (*27*). Such noise can arise from random fluctuations intrinsic to the system itself or from physiologically irrelevant environmental stimuli. Thus, a critical element in circuits with positive feedback is an associated damping process that prevents inadvertent activation. A simple "sink" or "filter" that blocks weak signals so that they do not reach a level capable of amplification is the simplest form of control. (In a biological setting, a constitutively functional phosphatase that inactivates a stochastically activated kinase is one possible example). However, the price for this type of control is an overall blunting of the sensitivity of the system, because a fixed amount of input is always silenced.

A negative-feedback circuit provides a more flexible form of control. Most of the time, when input is minimal, this repressive system is largely inactive. When either noise or low-level signals arise, they are countered

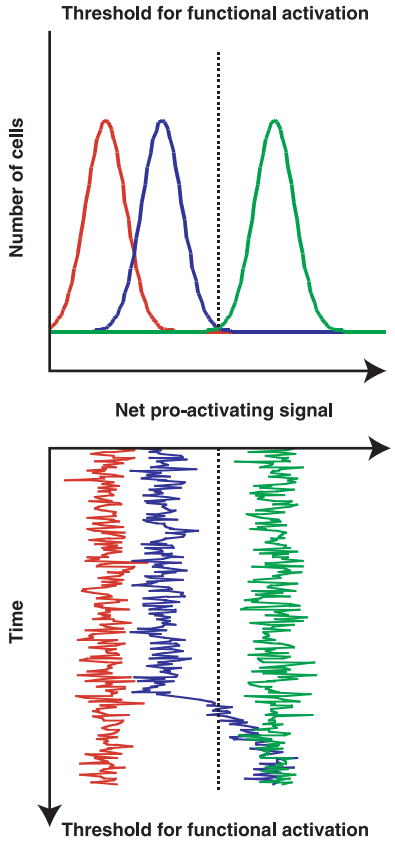

Fig. 1. The upper graph shows the distribution of autoantigen-specific T cells reaching particular levels of net activating signal (TCR and costimulatory input balanced by regulatory cell and cytokine inhibition) in a normal individual (red), an autoimmune disease–prone individual (blue), and an individual with disease (green). The threshold for functional activation is also shown. The lower graph illustrates the stochastic fluctuation over time in "net activation energy" of an individual naïve T cell in a normal individual, in a disease-prone individual, and among activated cells in a disease-prone individual, plotted against the threshold necessary for clonal expansion and effector differentiation. Activation energy varies with time owing to changes in intrinsic TCR and cosignaling capacity, availability of costimulatory input, ligand density, and inhibitory input. In the normal individual, it does not reach the threshold and no response occurs. In the disease-prone individual, the cell exceeds the threshold and responds. Once a cell becomes activated, it changes its capacity for signaling and response so that available activating inputs regularly exceed inhibitory ones, and the expanding cell population now experiences signals that consistently exceed the activation barrier.

by induction of a rising, proportional level of negative regulatory control. In the context of signaling, this could involve a phosphatase whose activity varies in response to the level of kinase induction. A major difference between tonic filters and negative-feedback circuits is the time delay inherent in the latter, which permits the input signal to mediate some effect before the effect of negative feedback is manifest (*28*). The delay might be of little consequence if the rate of product accumulation is so slow that it is blunted before reaching the threshold level necessary for any downstream effects. On the other hand, this property can be used to allow low-level, short-term responses followed by a return to baseline, or in a more sophisticated manner, to funnel inputs along one, but not another, pathway because of differing requirements for the concentration of a second messenger. This key role of timing in discriminatory signal processing will be expanded on below.

Another mechanism that suppresses inadvertent stimuli while facilitating sensitivity involves circuits with the property of ultrasensitivity. Reactions showing standard Michaelis-Menten properties go from 10 to 90% completion with an 81-fold change in the concentration of the variable component. If traversing this 10 to 90% interval requires a smaller change in input concentration, the system is said to be ultrasensitive (*29*), a property that often arises from cooperativity in the behavior of the interacting components. The steepness of the dose-response curve reflects a parameter termed the Hill coefficient, which at high values ($>\sim5$) provides the system with nearly digital switch-like effects (*30*). A valuable property of certain forms of ultrasensitivity is that the response is negligible at low input levels (Fig. 2). Thus, small spikes of input are suppressed and do not undergo positive amplification. However, a real signal that slightly exceeds in magnitude these incidental stimuli instead produces a maximal signal from the system. An example of such an ultrasensitive pathway is the one leading to mitogen-activated protein kinase (MAPK, also known as ERK) activation, which shows a high Hill coefficient (*31*). Once more than a low level of receptor input occurs, a full ERK response takes place. This type of digital behavior is especially obvious when ultrasensitivity is combined with positive feedback (*31*), which may help explain why the behavior of individual lymphocytes is so switchlike, given the central role of ERK in controlling gene activation and cell division.

Biological examples of feedback control. The several elements discussed above can be combined to produce more complex circuits with enhanced capacity for controlling inadvertent activation while promoting sensitivity to "real signals." Competing feedback loops, amplification, and a possible role for ERK pathway ultrasensitivity can be seen in the circuitry controlling signaling within a single cell—for example, whether TCR binding induces a functional response to a particular peptide: MHC ligand [Supplemental note 2 and Web fig. 2 (*15*)]. Another example of these regulatory principles involves CD4 effector differentiation—i.e., the T helper cell 1 (T_H1) versus T_H2 choice (*5*). Upon initial activation by antigen and a minimal set of cosignals, naïve $CD4^+$ T lymphocytes make only a few cytokines, primarily IL-2 and IL-3. Within a day or two, however, these cells can synthesize and secrete a variety of potent effector cytokines. Under many circumstances, there is a clear divergence in the capacity of individual activated $CD4^+$ T cells to make interferon-γ (IFN-γ) versus IL-4 (*32*). A series of feedback loops are involved in promoting either the T_H1 or T_H2 phenotype. Substances from certain infectious agents can prime antigen-presenting DCs for IL-12 production, while also up-regulating CD40 and the CD28 ligands CD80 and CD86 (*33*). T cells whose receptors bind tightly to an adequate quantity of processed antigen presented by major histocompatibility complex (MHC) class II molecules on these DCs in the context of CD28 costimulation respond by synthesizing CD40L. This stimulates the primed $CD40^+$ DCs to produce high levels of bioactive IL-12 and further increase costimulatory molecule expression, leading to IFN-γ production by the T cells that in turn enhances their sensitivity to IL-12 [Web fig. 3 (*15*)] (*34*). T_H2 development uses a distinct set of regulatory loops that again combine positive-feedback effects (IL-4 promoting IL-4 gene activity) (*35*) with negative effects (GATA-3 suppressing IL-12 reactivity) (*36*). The ultrasensitivity that results from the organization of the ERK pathway has already been discussed—it is a small leap to think that the effects of the cascade of T cell–DC interactions might share this property [Supplemental note 3 (*15*)].

It is important to point out that the polarizing effect of the cytokine signals is not necessarily immediate or absolute—rather, the feedback may evoke a small change in the probability that a cell will adopt a particular phenotype in each round of division; when this is iterated repeatedly during clonal expansion, the biasing effect can become large although not absolute because of the statistical nature of the change in gene expression (*4*–*7*). Alternatively, the choice of cell fate may be entirely stochastic, and the polarizing effect of a cytokine may relate to its capacity to mediate proliferative amplification of cells already committed to a certain differentiation pathway (*16*).

Spatiotemporal Considerations

Kinetics. Feedback control is only one device used by the immune system to focus its activity along productive channels. The orchestration or timing of each step in a complex series of events also is an important aspect of regulation (*37*). The system is organized such that by requiring more than one input within a narrow window of opportunity, fidelity can be substantially improved and false-positive events greatly reduced. This notion of "kinetic proofreading" serves as a guiding paradigm for how a T cell distinguishes the quality of a TCR ligand (*38*). Recent models have focused on the relative rates of ligand dissociation from the TCR and of coreceptor interaction with an engaged TCR complex as it moves within the membrane (*37*) (Fig. 3). Peptide-MHC combinations that

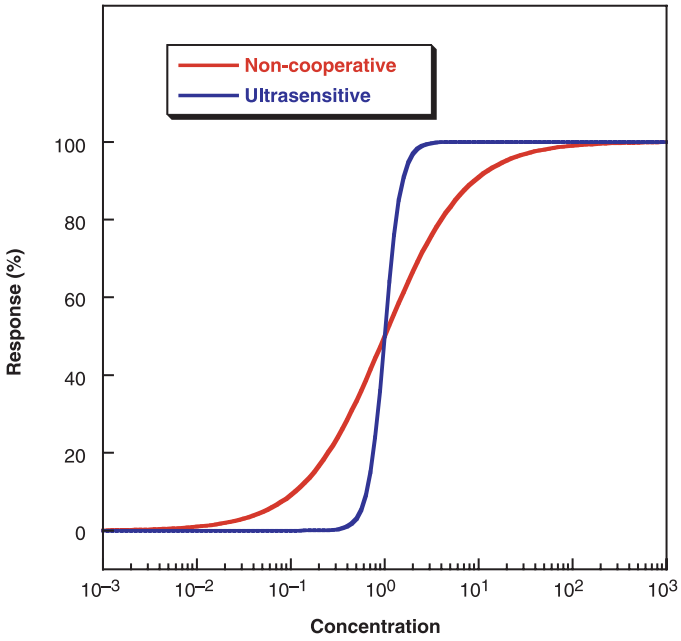

Fig. 2. Graph showing the suppression of low-amplitude inputs and nearly digital responses to inputs that exceed the low-amplitude threshold in a system showing ultrasensitivity, compared with a typical Michaelis-Menten (noncooperative reaction) dose-response curve.

dissociate rapidly from a TCR will not be capable of supporting the coordinate binding of a CD4 or CD8 coreceptor to both the TCR complex and the MHC ligand, because the two latter molecules are not spatially associated once dissociation takes place. However, a longer-lived interaction will allow a coreceptor to find the paired proteins and contribute to prolongation of the complex's lifetime, to augmentation of signaling through the additional Lck that is recruited to the TCR complex, and to protection of the complex from SHP-1 inactivation through a rapid rise in ERK activity (*33*). Thymic negative selection establishes a threshold that balances self-ligand off-rate with coreceptor access to any engaged TCR, on the basis of membrane surface area, coreceptor density, and receptor-coreceptor mobility (*39*). This prevents the TCR on nondeleted mature T cells from producing activating signals to self-ligands whose binding affinity is below this threshold.

Temporal control of responses can also arise from requirements for some minimal duration in one signaling state before the next input can be received. This mechanism helps limit inadvertent enhancement of low-amplitude inputs. It can also help direct a cell into distinct states of differentiation, depending on which of a series of hierarchical thresholds corresponding to signal duration is exceeded (*40*). One example involves the "ping-pong" interaction of T cells and DCs as discussed above. The first step in the process must be sufficiently long-lasting for the T cell to produce CD40L while still bound to the antigen-bearing, preactivated DC. Another example involves the commitment of $CD4^+CD8^+$ immature thymocytes to the CD4 versus CD8 lineages (*41*).

Location. Spatial organization plays a complementary role to temporal aspects of immune regulation. Early in the response, individual lymphocytes, even those with equivalent antigen specificity, meet depots of antigen in discrete sites too far apart for locally secreted mediators to readily influence lymphocytes at other antigen foci. For this reason, the effects of feedback amplification and control are most often local, not global. This dispersed nature of clonal immune responses prevents an inappropriate response from dominating the entire system. The result is greater diversity in the overall response than would be expected from in vitro studies that involve conditions resembling a "well-stirred chamber," avoiding the pathology produced by hyperpolarized T cell responses (*42*). An important consequence of this spatial aspect of immunity is that modeling immune responses requires consideration of the state of the local environment at the level of interacting cells.

Diversity in the immune response due to heterogeneity in local conditions is also of importance in considering the effects of the self-amplifying interaction between a T cell and a DC, or between a T cell and B cell. If "all-or-none" triggering applied to every antigen-activated lymphocyte, it would be difficult to understand the origin of the variety of behaviors seen among these responding cells, such as differentiation into terminal effectors versus central memory cells (*43*), separation into cells destined to survive after expansion or to die (*44*), or differentiation into B cells with distinct immunoglobulin isotypes (*45*). The possibility that both switchlike behavior as well as variation function in the outcome of cell-cell interactions becomes reasonable if certain conditions prevail—if, for example, different lymphocytes show a distribution of signaling competence, if the DCs presenting a given antigen are functionally heterogeneous, and if the T cells show a broad quantitative distribution of cytokine production (*46*). For example, the T cells with the greatest signaling competence, meeting a DC with high amounts of processed antigen, primed

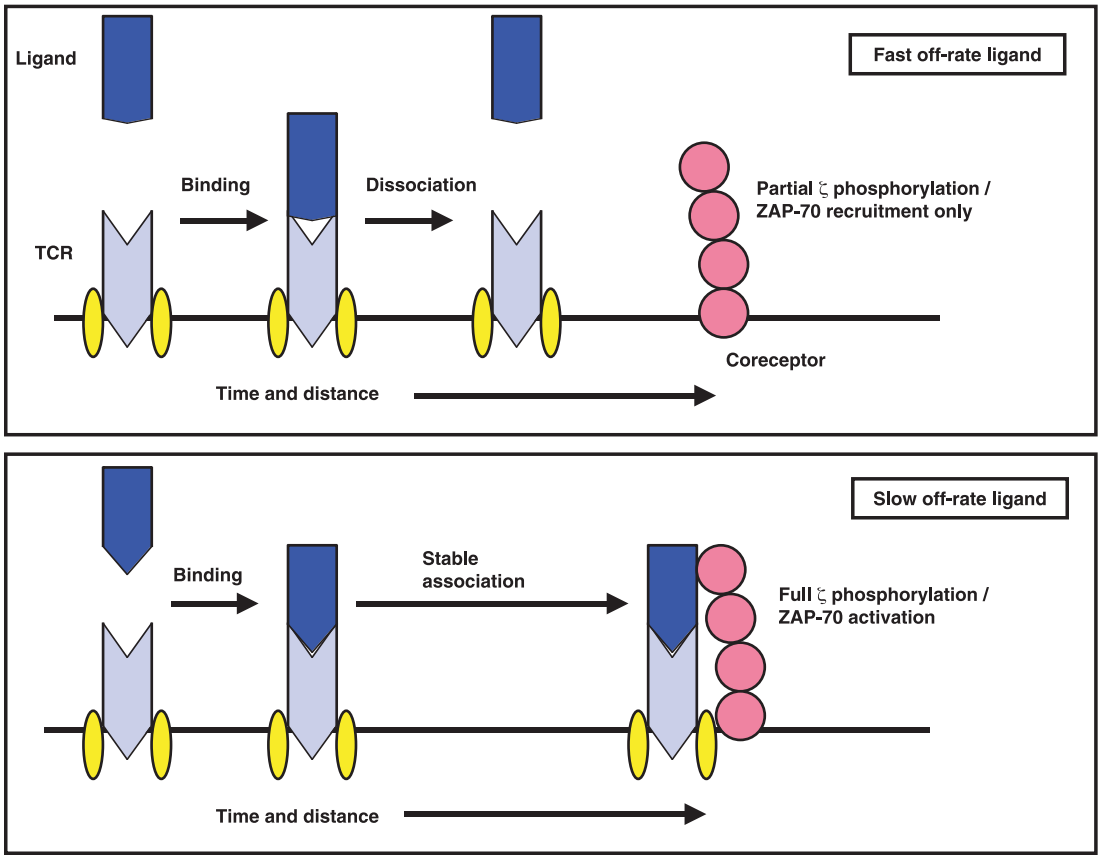

Fig. 3. Illustration of how differences in the rate of ligand-TCR dissociation can affect signal quality through a kinetic proofreading step involving interaction with spatially distinct coreceptors. Fast off-rate ligands can initiate signaling. However, they do not remain bound to the TCR for a time adequate for a coreceptor located at some distance on the membrane from the engaged complex to come into proximity and form a ternary complex. This aborts signaling at an incomplete stage, perhaps owing to the dominant SHP-1 negative-feedback effect (*33*). Ligands with slower dissociation rates not only initiate signaling but remain bound to the TCR for a sufficient time to come into contact with a coreceptor, forming a ternary complex that can more effectively activate ERK, blocking the SHP-1 negative-feedback loop, and extending the time available to the TCR complex to generate downstream signals for synapse formation and eventual gene activation. Thus, the expression level of coreceptor (which sets the mean distance to an engaged TCR and hence the time it takes to find such a binary complex) dictates the minimal quality (off-rate) of a TCR-ligand pair that will produce activating signals. This threshold is set by thymic selection so that self-ligands do not exceed this limit, but foreign ligands will do so for the appropriate TCR.

by infectious signals to express CD40, and ready for an IL-12 response to CD40 ligation will be expected to show strict development into fully polarized T_H1 cells. These cells may constitute terminal effectors primarily resident in nonlymphoid tissues. Other T cells of the same specificity, but with lower signaling competence, or having encountered DCs not primed to express CD40 or IL-12, will be unable to take advantage of the self-amplifying polarization cycle. These cells might instead differentiate into primed memory cells that retain naïve cell-like expression of CCR7 and a capacity for IL-2 production (*43*). By this mechanism, some cells can take advantage of the full set of amplifying mechanisms built into the system, developing into highly active effectors needed to fight the ongoing infection, whereas others are shunted into memory pools that serve as a reservoir for continuing or subsequent responses to the same antigen. These decisions are made in a statistical fashion, but given a suitably robust response, the overall behavior of the system is relatively predictable in its proportional production of the various cell types.

Stochastic Gene Expression

With the exception of antigen receptor genes that show allelic exclusion, most genes are thought to be expressed from both alleles when the differentiated state of the cell supports transcription from that locus. Data demonstrating that some cytokine genes are often expressed in a monoallelic manner during any given activation cycle of effector T cells were thus very unexpected (*47*). Why should the system behave in this manner? One possibility is to limit the level of potentially harmful mediators, although this would result in only a twofold decrease compared with biallelic activation. Recent modeling of the condition known as haploinsufficiency (*48*) suggests an alternative explanation. Gene transcription is episodic, resulting in oscillations in the concentrations of end product. When both alleles are active, temporally noncoordinate transcription tends to smooth out these ups and downs. For proteins of limited stability, the mean level when only one allele can be transcribed is 50% of that seen when both alleles are potentially active, but the variation around the mean is much greater. Depending on the relative rates of transcription, mRNA degradation, and protein degradation, a cell transcribing from only one allele may experience intervals when the protein in question falls below 20% of the maximum seen with biallelic expression (Fig. 4). It is not difficult to imagine that such low levels of a molecule could interrupt a circuit in which it is a crucial feedback component. In the case of $CD4^+$ T cells, this could limit polarization of an expanding T cell clone, or leave cells within the clone susceptible to the influence of counterregulatory cytokines. Thus, monoallelic expression of cytokine genes might not only limit potential immunopathology, but might also decrease the drive of $CD4^+$ T cell responses toward full polarization. This would afford the system an opportunity for enhanced regulation by opposing cytokines and provide the evolving response with the potential for greater functional diversity.

Conclusion

In fulfilling its essential role in host defense, the immune system balances robust responses to foreign antigens with tight control of overt self-reactivity, the maintenance of diverse repertoires of antigen-reactive lymphocytes with effective tolerance, and cell-mediated reactions with humoral antibody production. Most lymphocytes generated in response to infection die, but some remain part of the peripheral lymphoid pool for much of the lifetime of the host. Adaptive immune responses are crucial for adequate protection, but innate immune mechanisms are needed to blunt pathogen growth, to allow time for the adaptive system to respond, and to direct the latter to the appropriate class of effector function (*49*).

This review has emphasized that these complex, balanced behaviors arise from the repetitive use of a small set of basic mechanisms that amplify weak signals to generate responses of sufficient magnitude to deal with rapidly reproducing pathogens, that increase the discrimination between noise and useful information, that prevent the overproduction of potentially dangerous cells and mediators, and that diversify responses even in the face of positive-feedback processes that tend to homogenize output. It has also stressed that rather than arising from globally connected interactions that instruct the behavior of the entire cohort of responding cells as a unit, immune responses represent a statistical summing of individual cellular responses given a bias by signals that modify the likelihood of a particular behavior by a responding cell, further amplified by massive cell replication and sometimes nearly equally massive cell death.

All of this argues that in addition to the more typical intuitive interpretation of immunological data, it is now time to add the power of mathematics, systems analysis, and quantitative cell-based modeling. These latter approaches have already yielded valuable results that are helping to guide further biological experimentation on topics such as signaling in *Drosophila* development (*50*). Analysis of systems as seemingly diverse as the WWW (*51*) and yeast metabolic pathways (*52*) is beginning to document the power function organization of connections in these complex networks. Such organization conveys an "error resistant"/"attack prone" character on the system, which in turn helps to explain why some gene deletions show little biological effect (they are of low connectivity) and others are lethal (high connectivity) (*52*, *53*). By designing immunological experiments to yield the dynamic, spatially resolved, quantitative information that is needed to properly describe functional components and pathways in the immune system in a manner suitable for "reverse engineering," we will find ourselves better able to predict the system's operation in health and disease. This predictability will also permit us to design

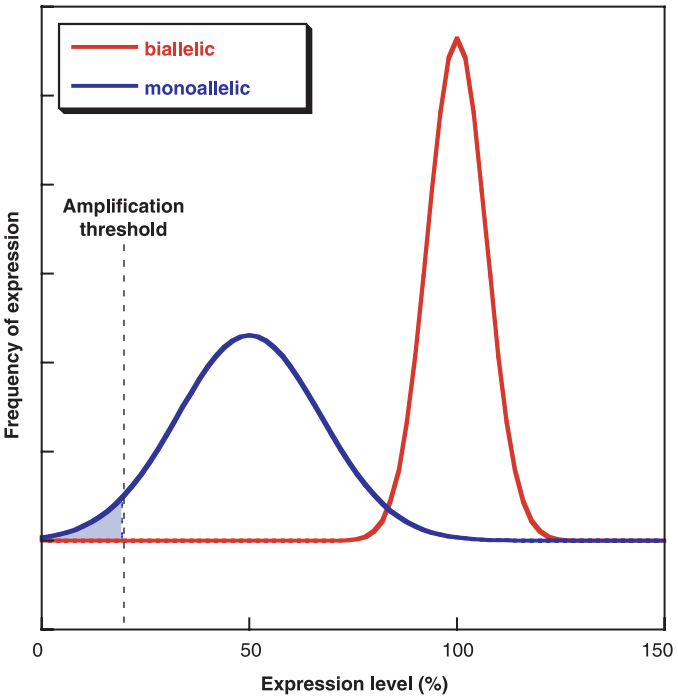

Fig. 4. Illustration of the increased variation around the mean seen in the level of a gene product when one versus two genes are being transcribed, and the likelihood that with only one allele active, the product level will drop below a threshold (20% expression) necessary to sustain a particular cellular function or state. This is a form of "haploinsufficiency," and the limitation of polarizing positive-feedback effects through this mechanism may be one reason that cytokine genes show monoallelic expression.

rational interventions that augment, depress, or deviate responses in ways that promote human health.

References and Notes

1. M. Cohn, R. E. Langman, *Immunol. Rev.* **115**, 11 (1990).
2. This description of the immune system is a paraphrase of the classic children's tale "Goldilocks and the Three Bears." Upon entering the empty bear's home in the forest, Goldilocks found the porridge on the table "Too hot," "Too cold," and "Just right."
3. R. P. Bucy *et al.*, *J. Exp. Med.* **180**, 1251 (1994); Y. Itoh, R. N. Germain, *J. Exp. Med.* **186**, 757 (1997).
4. A. Kelso, P. Groves, A. B. Troutt, K. Francis, *Eur. J. Immunol.* **25**, 1168 (1995).
5. S. L. Reiner, R. A. Seder, *Immunity* **11**, 1 (1999).
6. R. Callard, A. J. George, J. Stark, *Immunity* **11**, 507 (1999).
7. A. V. Gett, P. D. Hodgkin, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9488 (1998).
8. A. V. Gett, P. D. Hodgkin, *Nature Immunol.* **1**, 239 (2000).
9. W. E. Paul, Ed., *Fundamental Immunology*, (Lippincott-Raven, New York, ed. 4, 1999); C. A. Janeway Jr., P. Travers, M. Walport, J. D. Capra, Eds., *Immunobiology: The Immune System in Health and Disease* (Garland, New York, 1999).
10. D. L. Hartl, A. G. Clark, *Principles of Population Genetics* (Sinauer, Sunderland, MA, ed. 2, 1989).
11. P. Bretscher, M. Cohn, *Science* **169**, 1042 (1970).
12. R. H. Schwartz, *Science* **248**, 1349 (1990).
13. C. B. Thompson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 1333 (1989).
14. D. A. Cantrell, K. A. Smith, *Science* **224**, 1312 (1984).
15. Supplementary data are available on *Science* Online at www.sciencemag.org/cgi/content/full/293/5528/240/DC1.
16. A. C. Mullen *et al.*, *Science* **292**, 1907 (2001).
17. D. A. Carson, *FASEB J.* **6**, 2800 (1992).
18. S. Sakaguchi, *Cell* **101**, 455 (2000).
19. J. M. Curtsinger, D. C. Lins, M. F. Mescher, *J. Immunol.* **160**, 3236 (1998); M. Tsukamoto, R. N. Germain, unpublished observations.
20. M. Croft, L. M. Bradley, S. L. Swain, *J. Immunol.* **152**, 2675 (1994); M. Pihlgren, P. M. Dubois, M. Tomkowiak, T. Sjogren, J. Marvel, *J. Exp. Med.* **184**, 2141 (1996); P. R. Rogers, C. Dubey, S. L. Swain, *J. Immunol.* **164**, 2338 (2000).
21. P. Matzinger, *Annu. Rev. Immunol.* **12**, 991 (1994).
22. P. V. Lehmann, T. Forsthuber, A. Miller, E. E. Sercarz, *Nature* **358**, 155 (1992).
23. C. B. Thompson, J. P. Allison, *Immunity* **7**, 445 (1997).
24. H. Nishimura, M. Nose, H. Hiai, N. Minato, T. Honjo, *Immunity* **11**, 141 (1999).
25. J. D. Katz, B. Wang, K. Haskins, C. Benoist, D. Mathis, *Cell* **74**, 1089 (1993).
26. N. Barkai, S. Leibler, *Nature* **387**, 913 (1997).
27. M. Freeman, *Nature* **408**, 313 (2000).
28. Z. Grossman, W. E. Paul, *Semin. Immunol.* **12**, 197 (2000).
29. D. C. LaPorte, K. Walsh, D. E. Koshland Jr., *J. Biol. Chem.* **259**, 14068 (1984).
30. C. Y. Huang, J. E. Ferrell Jr., *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10078. (1996).
31. J. E. Ferrell Jr., E. M. Machleder, *Science* **280**, 895 (1998).
32. W. E. Paul, R. A. Seder, *Cell* **76**, 241 (1994).
33. J. Aliberti *et al.*, *Nature Immunol.* **1**, 83 (2000); O. Schulz *et al.*, *Immunity* **13**, 453 (2000); H. Hemmi *et al.*, *Nature* **408**, 740 (2000).
34. C. A. Wenner, M. L. Guler, S. E. Macatonia, A. O'Garra, K. M. Murphy, *J. Immunol.* **156**, 1442 (1996).
35. R. A. Seder, W. E. Paul, M. M. Davis, B. Fazekas de St. Groth, *J. Exp. Med.* **176**, 1091 (1992).
36. W. Ouyang *et al.*, *Immunity* **9**, 745 (1998).
37. R. N. Germain, I. Stefanova, *Annu. Rev. Immunol.* **17**, 467 (1999).
38. T. W. McKeithan, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5042 (1995).
39. S. C. Jameson, K. A. Hogquist, M. J. Bevan, *Nature* **369**, 750 (1994).
40. G. Iezzi, K. Karjalainen, A. Lanzavecchia, *Immunity* **8**, 89 (1998).
41. K. Yasutomo, C. Doyle, L. Miele, C. Fuchs, R. N. Germain, *Nature* **404**, 506 (2000).
42. K. F. Hoffmann, A. W. Cheever, T. A. Wynn, *J. Immunol.* **164**, 6406 (2000).
43. F. Sallusto, D. Lenig, R. Forster, M. Lipp, A. Lanzavecchia, *Nature* **401**, 708 (1999).
44. J. K. Whitmire, K. Murali-Krishna, J. Altman, R. Ahmed, *Philos. Trans. R. Soc. London Ser. B* **355**, 373 (2000); J. Jacob, D. Baltimore, *Nature* **399**, 593 (1999).
45. C. M. Snapper, F. D. Finkelman, W. E. Paul, *Immunol. Rev.* **102**, 51 (1988).
46. A. Lanzavecchia, F. Sallusto, *Science* **290**, 92 (2000).
47. I. Riviere, M. J. Sunshine, D. R. Littman, *Immunity* **9**, 217 (1998); M. Bix, R. M. Locksley, *Science* **281**, 1352 (1998); J. Hu-Li *et al.*, *Immunity* **14**, 1 (2001).
48. D. L. Cook, A. N. Gerber, S. J. Tapscott, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15641 (1998).
49. C. A. Biron, *Semin. Immunol.* **10**, 383 (1998).
50. G. von Dassow, E. Meir, E. M. Munro, G. M. Odell, *Nature* **406**, 188 (2000).
51. S. H. Strogatz, *Nature* **410**, 268 (2001).
52. H. Jeong, S. P. Mason, A. L. Barabasi, Z. N. Oltvai, *Nature* **411**, 41 (2001).
53. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A. L. Barabasi, *Nature* **407**, 651 (2000).

54. I regret that it was not possible to cite all relevant references owing to space limitations. I thank I. Stefanova, J. Dorfman, J. Delon, G. Bonnet, and M. Meier-Schellerscheim for helpful comments on the manuscript, A. Kirk for pointing out the similarities between the scenario for prophylaxis of autoimmunity and current research in transplantation combining lymphoid depletion and low-dose immunosuppression, and G. Bonnet for help with the figures.

VIEWPOINT

T Cell Death and Memory

Jonathan Sprent* and David F. Tough†

In typical immune responses, contact with antigen causes naïve T cells to proliferate and differentiate into effector cells. After the pathogen is destroyed, most effector T cells are eliminated—thereby preserving the primary T cell repertoire—but some cells survive and form long-lived memory cells. During each stage of this process, the life or death fate of T cells is strictly regulated.

Immune responses leading to rejection of infectious agents usually culminate in a state of specific T and B cell memory where secondary responses are more vigorous and effective than primary responses (*1*–*6*). Generation of memory T and B cells is the end result of a highly destructive process in which most of the responding lymphocytes are rapidly eliminated, and only a small proportion survive to become long-lived memory cells. This article reviews the life or death decision-making involved in the formation of memory T cells, as well as the role of certain cytokines in keeping these cells alive.

Longevity of naïve T cells. Naïve T cells are long-lived resting cells that reside in the recirculating lymphocyte pool and migrate continuously from blood to lymph through specialized T cell zones in the secondary lymphoid tissues, the spleen, lymph nodes (LNs), and Peyer's patches (*7*). The survival of naïve T cells requires continuous contact with self peptides bound to major histocompatibility complex (MHC) molecules combined with exposure to a cytokine, interleukin 7 (IL-7) (*6*). In consort, these two ligands are presumed to induce a form of low-level signaling that is sufficient to keep the cells alive but does not induce them to enter the cell cycle.

Life and death during the primary response. Primary T cell responses are initiated in secondary lymphoid organs by mature antigen-presenting cells, i.e., dendritic cells (DCs) (*8*). Recognition of immunogenic peptides bound to cell-surface MHC molecules on DCs in the T cell zone causes selective sequestration ("trapping") of antigen-specific recirculating T cells entering lymphoid tissues from the blood (*9*); the trapped cells are then induced to proliferate.

Because infectious agents often replicate at a prodigious rate, primary immune responses are geared to be as intense as possible. Division of antigen-reactive T cells during the height of the immune response is very rapid (three to four divisions per day for $CD8^+$ cells) and leads to >1000-fold expansion of the responding cells within a few days (*10*). After differentiating into effector cells, the progeny of the responding cells reenter the circulation through efferent lymph and disseminate throughout the body (*11*–*13*). By means of expression of new cell surface–homing molecules, the effector cells acquire the capacity to penetrate capillary blood ves-

Department of Immunology, IMM4, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.

*To whom correspondence should be addressed: jsprent@scripps.edu

†Present address: The Edward Jenner Institute for Vaccine Research, Compton, Newbury, Berkshire RG20 7NN, UK.