The far reaches of enzymology

Jesse G Zalatan & Daniel Herschlag

The scope of enzymology has expanded rapidly over the last century, from an early focus on the chemical and catalytic mechanisms of individual enzymes to more recent efforts to understand enzyme action in the context of dynamic, functional biological systems consisting of many interacting enzymes and proteins. Continuing progress in probing the link between molecular structure and function now promises to pave the way for a deeper understanding of the evolution and behavior of the complex biological systems that govern cellular behavior.

Remarkable progress in chemistry and biology over the last century has provided deep insights into two fundamental questions: how do molecular components give rise to functioning biological systems, and how do complex biological systems evolve? We now have, in mechanistic detail, a considerable understanding of how biological macromolecules control growth, metabolism, signaling and other biological processes at the level of individual components, while sequence and structural comparisons have revealed the molecular-level changes that distinguish evolutionarily related macromolecules. At the heart of these biological processes are chemical reactions that are coordinated in space and time to produce a living system. To fully describe and understand these processes, it is necessary to go beyond the identification of the components and their interacting partners. Now that more complex aspects of biology are becoming accessible to investigations at the molecular and chemical level, the powerful approaches of mechanistic enzymology will enable the incisive dissection and integration of biological systems that are necessary for deep comprehension and ultimately rewiring and reconstruction of these systems and their regulation.

The development of enzymology

Applications of chemical approaches to biology in the twentieth century led to the birth of modern enzymology. The discovery of vitamins and then of metabolic pathways and products in the early twentieth century spurred interest in understanding the underlying chemical transformations that produce these molecules. Around mid-century, this knowledge and the chemical tools and assays available, coupled with the emerging area of protein purification, led to an explosion of information about biological reactions and the enzymes that catalyze them. The stage was set for the emergence of enzymology as a central science of biochemistry.

Though chemistry was the foundation for revealing the inner workings of biology at the time, biology spurred new advances in chemistry, with a plethora of new compounds and new reactions to study, efforts to mimic biological processes, and the overarching question of how enzymes could achieve their enormous rate enhancements and exquisite specificities. Many of the 'new' biological reactions, such as those involving phosphate esters and anhydrides, were not thoroughly covered in the existing chemical literature, and the field of bioorganic chemistry was born (**Fig. 1**)[1]. Bioorganic chemistry used traditional and emerging tools of physical organic chemistry and led to deep understanding of many reaction classes. Nevertheless, it later emerged that enzyme-catalyzed reactions do not always follow the same mechanism as uncatalyzed reactions, so the combination of studying reactions with and without their enzyme catalyst has been critical. Analogously, as we will see below, the study of biological pathways and systems requires *in vitro* reconstitution of components via the tools of enzymology, but also requires *in vivo* tests to confirm the predictions made by analyzing and perturbing *in vitro* systems.

Obtaining enzymes in pure form was an essential prerequisite for understanding the underlying chemical processes in biological systems, well before the advent of modern methods for overexpression and affinity tagging[2]. In addition to the most basic task of identifying distinct enzymes that catalyze distinct reactions (an endeavor that continues in the genomic era), enzymes were shown to exhibit high specificity, enforce stereochemistry, react via covalent catalysis, utilize general acid/base catalysis in proton transfer steps and manipulate the pK_a (protonation state) of active site residues[3–8]. What may be unimaginable to the contemporary scientist is that these physical models for the action of enzymes were developed in the absence of direct structural models. How would modern-day protein sciences proceed if we were suddenly blinded to the general and specific properties and features of proteins and their active sites? Today, as the frontiers of biochemistry encompass complex and dynamic macromolecular assemblies that carry out complex biological functions but are difficult to visualize structurally, the field is in need of modern-day enzymologists and biologists to make analogous conceptual advances into the workings of complex systems.

A key insight in the transformation of enzymology into a rigorous science was the realization that basic chemical principles could be applied to enzyme action. As we understand its history, the earliest recognition that binding to a catalyst could lower the energy barrier for a chemical reaction came from Polanyi in 1921 (refs. 9,10). More than 20 years later, Pauling articulated this idea in terms of preferential transition-state recognition to account for enzyme activity:

Jesse G. Zalatan is in the Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California, USA, and Daniel Herschlag is in the Department of Biochemistry, Stanford University, Stanford, California, USA.
e-mail: zalatan@cmp.ucsf.edu or herschla@stanford.edu

"...the only reasonable picture of the catalytic activity of enzymes is that which involves an active region of the surface of the enzyme which is closely complementary in structure not to the substrate molecule itself, but rather to the substrate molecule in a strained configuration, corresponding to the "activated complex" for the reaction..."[11].

Another 20 years later, Jencks began to put down his ideas on enzymatic catalysis in his classic 1969 book *Catalysis in Chemistry and Enzymology*[3]. Here Jencks systematized the current understanding of catalytic mechanisms, recognizing that:

"The discovery that certain enzymes react chemically with their substrates to form covalently bonded enzyme-substrate intermediates has done more than anything else to dispel mysterious mechanisms and vitalistic theories of enzyme action because it suggests that the mechanism of enzyme action is not fundamentally different from that of any other chemical reaction... In spite of the many important advances which have been made, it is important to keep firmly in mind that we do not in any case have a detailed or quantitative understanding of the mechanism and driving forces which account for the enormous specific rate accelerations which are brought about by enzymes...". [3]

Most powerfully, Jencks emphasized a fundamental connection between binding energy and enzymatic catalysis, and subsequent writings from Jencks further developed this connection and its implications throughout biology[12–14]. Most basically, these writings provided a clear explanation for how binding interactions remote from the site of chemical transformation could be used for preferential transition-state stabilization[13].

A related conceptual problem at the time centered around the suggestion that an effective molarity of ~55 M, corresponding to the concentration of surrounding solvent, would be roughly the maximal achievable rate advantage from binding and positioning two substrates together, whereas model bioorganic systems were known to exhibit "effective molarities" orders of magnitude greater than this value. Several new physical models were proposed to account for this discrepancy, most notably the "orbital steering" model, which suggested that the shape of electron orbitals leads to a sharp dependence of chemical reactivity on the relative orientation of reacting groups[15]. However, a systematic accounting of the loss of entropy in going from a bimolecular to a unimolecular state was sufficient to explain the discrepancy[12], whereas orbital steering was shown to contravene known properties of molecular bonding[16]. These findings indicated that existing physical and chemical principles were sufficient to explain the observed rate enhancements. Although much remains uncertain in the application of straightforward concepts of substrate conformational entropy to the solution phase (where there is thermodynamic coupling to solvent molecules and to the protein scaffold), here we emphasize that basic chemical principles provide important, necessary and sufficient insights into the behavior of biological systems.

As modern chemists and biologists tackle ever more dauntingly complex macromolecular assemblies and biochemical networks, it is increasingly important that the fundamental physical and chemical principles are kept in mind and brought to bear on these problems. Correspondingly, imparting this accumulated knowledge of physical organic chemistry, enzymology and physical chemistry to the next generation of chemists and the emerging cadre of chemical biologists will be critical.

Enzymology in complex reaction systems

Although enzymology was initially focused on purifying and investigating individual enzymes, it was recognized at the time that biology constitutes macromolecules working in concert and that biological understanding would require reckoning with systems of enzymes. Arne Tiselius, who won the Nobel Prize in Chemistry in 1948 for advances in electrophoresis, stated in his Nobel lecture:

"Throughout the course of biochemistry—quite naturally—principal interest has been directed to the chemical entities that have been isolated from biological materials, but there is little doubt that in the future much closer attention must be paid also to the existence of specific complexes in the living organism. It is the chain of events with linked reactions which is the key note of life's processes and this interrelationship or linkage can perhaps best be conceived as the result of an association of substances with different specific functions"[17].

Despite this early recognition of the importance of such studies, progress in reconstituting complex systems has been slow, and there are likely several underlying reasons. First, purification is difficult, and was much more so before cloning, overexpression, tagging and rapid chromatography methods became routine. Second, and pertinent to the situation today, scientists with expertise in enzymology and reaction mechanisms were less familiar with biological systems, their components and the critical surrounding questions. Nevertheless, researchers who have rigorously embraced complex biological systems at a molecular and quantitative level have made remarkable progress.

DNA polymerase is a classic example of an enzyme that functions in concert with a host of associated factors. To understand how this

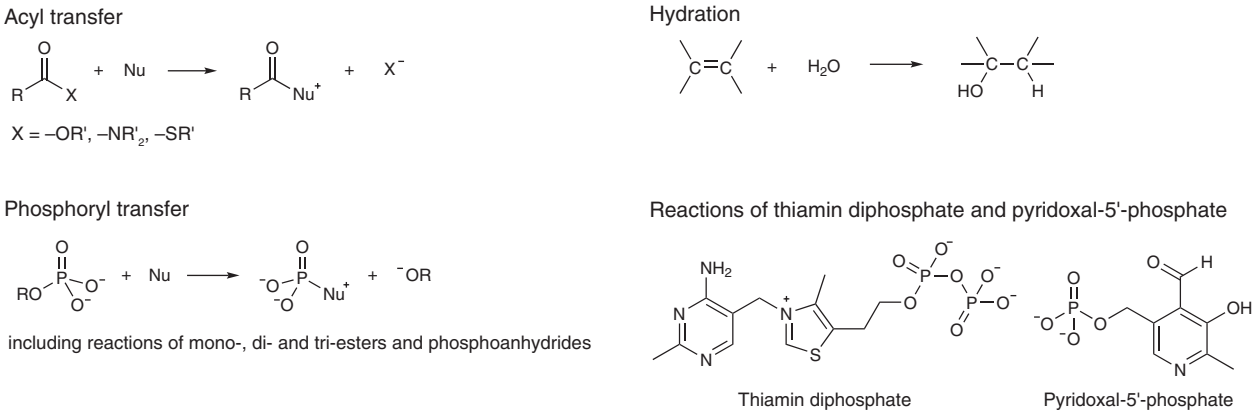

Figure 1 Basic biological reactions. The reactions depicted are selected from Bruice and Benkovic's seminal 1966 text *Bioorganic Mechanisms*[1], which also covers then-current topics in the chemistry of NADH, NADPH, folic acid and biotin.

machinery works together to replicate DNA, Benkovic and co-workers chose the simplest machinery, that from a bacteriophage, as a model system. Over the past three decades, this detailed work has led to remarkable insights into the individual reaction steps involved in the workings and assembly of the T4 replisome[18]. Sacrificing the relative simplicity of quantitative mechanistic dissections with the minimal T4 replisome, O'Donnell and co-workers set out to reconstitute all of the factors involved in the more complex replication of a bacterial genome. These studies identified a processivity factor that could dissociate from linear but not circular DNA, leading to the prediction that the protein formed a 'sliding clamp' around DNA—an insight later confirmed by an X-ray structure[19].

Indeed, pioneering efforts at structural determination of enzymes and proteins associated with DNA replication by Thomas Steitz and contributions by many others have provided numerous insights and have made 'real' the ephemeral nanoworld of biology. The progress toward understanding basic features of replication from these and other labs has been astounding, yet it is also clear that the problem is still much more complex. DNA replication is coupled to repair and transcription and is controlled by cell cycle factors. From a functional perspective, understanding how these processes are integrated requires quantitative studies that dissect individual steps and reveal the connections between the various components. From a structural perspective, one of the greatest current challenges is visualizing the functional complexes, including many that are transient or membrane-associated, that integrate these processes. Here we emphasize that enzymological characterization can be crucial in such endeavors by defining the conditions for formation of various complexes, and that, once structural information is obtained, enzymological characterization is essential to define the actual workings of such complexes. Further, as emphasized immediately below, determining whether the reconstituted system is sufficient to account for actual biological function is a necessary additional step.

Identifying and characterizing the essential components of functional, dynamic biological complexes by biochemical reconstitution coupled with quantitative kinetic and thermodynamic measurements and comparison to cellular function led to the discovery of GTPase activating proteins (GAPs). This history is not widely appreciated, although the significance of GAPs for G protein signaling is well established and routinely taught in introductory biochemistry courses. In the late 1980s, G proteins were known to cycle between an active GTP-bound state and an inactive GDP-bound state, and the active-to-inactive transition was thought to be mediated by the intrinsic GTPase activity of the G proteins. However, oncogenic G protein mutants were constitutively active *in vivo* but did not always have slower rates of GTPase activity. Thus, a factor that would affect turnover was sought, and quantitative *in vivo* and *in vitro* experiments measuring nucleotide binding and hydrolysis rates uncovered a soluble protein factor, the GAP, that accelerated the rate of wild-type GTPase activity but had no effect on the GTPase activity of the oncogenic mutants[20]. These results implied that additional proteins played an important role in the regulation of G protein signaling and significantly expanded the scope of the field. This fundamental discovery about the behavior of G protein signaling networks was prompted by a quantitative discrepancy between *in vivo* and *in vitro* assays of enzyme activity.

More recently, quantitative experiments in reconstituted systems have provided important insights into the role of Arp2/3 in actin filament nucleation and actin-based cell motility[21]. Another recent success was the reconstitution of a three-component circadian oscillator from cyanobacteria and the derivation of a quantitative framework for the system[22]. Many more systems await such dissection and integration with *in vivo* observations, and we expect that understanding of biological systems will rapidly increase as these mechanistic frameworks are developed.

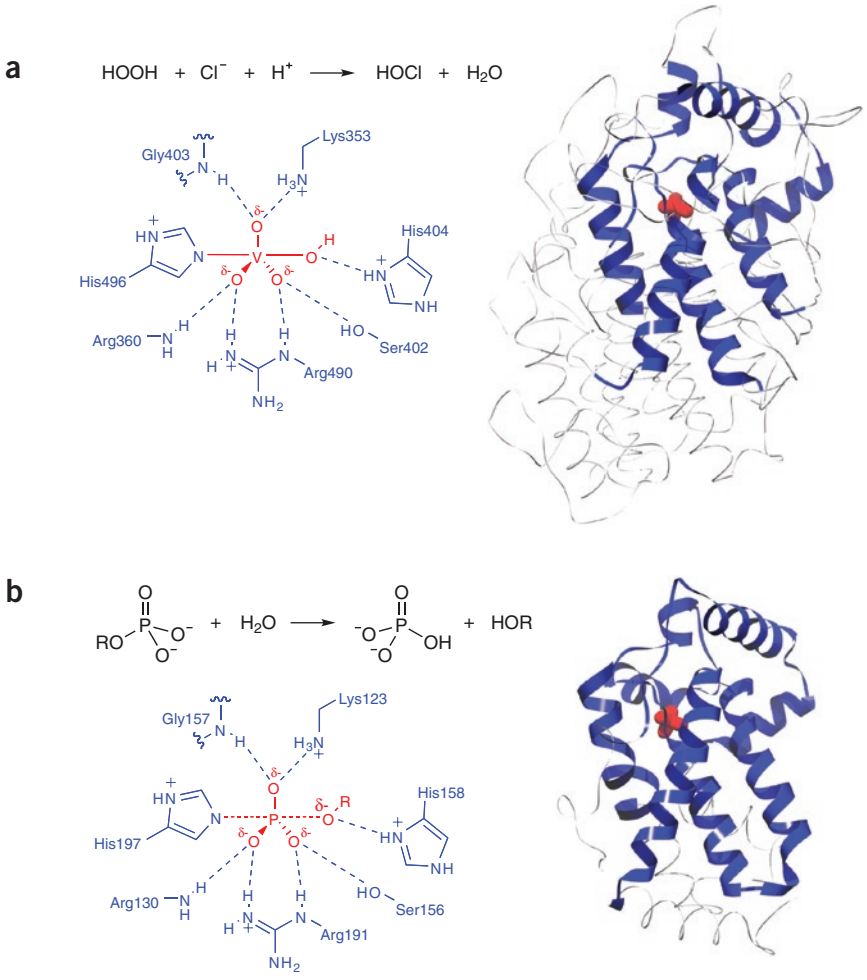

Figure 2 Divergent reactivity in evolutionarily related enzymes. (**a**,**b**) Comparison of vanadium-dependent chloroperoxidase (Protein Data Bank ID 1IDQ) (**a**) and PAP2 acid phosphatase (Protein Data Bank ID 2AKC) (**b**). The ribbon diagrams show overall structural homology between the two proteins (thick ribbons) and the crystallographically observed active site ligands (vanadate and tungstate, respectively) in red. The active site schematics show the conserved active site residues (blue) interacting with vanadate (**a**) and a model for the transition state for phosphoryl transfer (**b**). The overall reaction catalyzed by each enzyme is shown above each structure. Molecular graphics images were produced using the UCSF Chimera package[30].

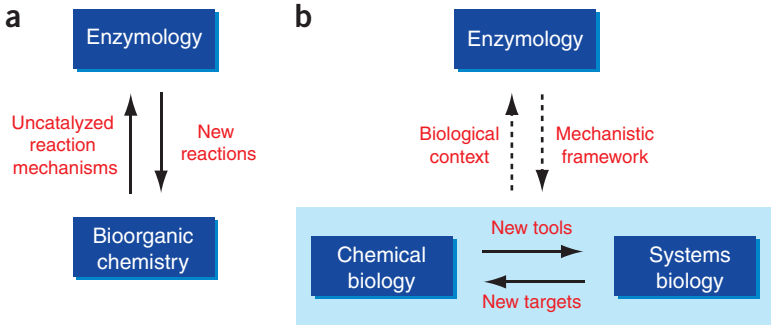

Figure 3 Enzymology past, present and future. (**a**) In the mid-twentieth century, enzymology provided bioorganic chemists with the motivation to study new, previously uncharacterized reactions. These studies provided the framework for comparisons of catalyzed and uncatalyzed reaction mechanisms. (**b**) At the beginning of the twenty-first century, the fields of chemical biology and systems biology are tightly integrated, with chemical biology providing tools to probe the behavior of biological systems, and systems biology spurring focus on new targets to interrogate chemically. The challenge for modern students of chemistry and biology is to integrate the insights provided by enzymology into a rigorous and deep mechanistic framework for the behavior of biological systems (dashed lines).

Enzymology and evolution

Fundamental questions in the evolution of biological systems can be addressed from the molecular perspective provided by enzymology. Indeed, this perspective is indispensable for connecting molecular changes at the level of nucleic acids and proteins to changes in the behavior of a biochemical system and for understanding the underlying physical and chemical origins of the change. Two key underpinnings of molecular evolution—the RNA world hypothesis and the catalytic promiscuity model for enzyme evolution—derive fundamental insights from enzymology. Below we highlight some exciting recent developments that have emerged from research motivated by these ideas.

RNA can act as both an information carrier and a functional macromolecule. The recognition nearly 30 years ago that RNA can catalyze phosphodiester bond formation implied that the core chemical functionality for an RNA world was possible, and much attention in the field has been focused on obtaining, via *in vitro* selection and molecular engineering, a ribozyme that can self-replicate[23]. Capping over 20 years of progress toward this goal, a self-replicating RNA ligase that assembles copies from oligonucleotide substrates was shown to carry out multiple sustained rounds of directed evolution[24]. Continued directed evolution may lead the way to bona fide self-replicating polymerases that can act on monomeric nucleotide precursors.

Early attention to the RNA world hypothesis led to experiments exploring the ability of RNA to bind a plethora of potential metabolites and catalyze a variety of types of reactions. Since these investigations, the field of RNA enzymology has matured, and now lessons from naturally occurring ribozymes can be applied to the selection and design of new RNA catalysts, or at least catalysts that are new to our current protein-dominated world. For example, what metabolites can be synthesized, and what polymers can be assembled? Can metabolic pathways necessary for early life be developed with an all-RNA cast of catalysts? And along the way, can we derive new generalizations about biological catalysis and new insights into functional RNAs and proteins[25]?

Just as an understanding of the chemical capabilities of RNA led to new ideas about early evolution, a chemical perspective on enzyme specificity has led to new ideas about the evolution of new enzyme activities. Although one of the classical hallmarks of enzymes is their exquisite specificity, we now recognize that many enzymes exhibit "catalytic promiscuity"—the ability of an enzyme to catalyze, at a low level, a reaction other than its cognate reaction[26,27]. Subsequent to a gene duplication event, enzymes that can promiscuously catalyze an alternative reaction may have a selective advantage for optimization of that activity. Promiscuity arises from the fact that enzyme functional groups optimized to catalyze one reaction can often stabilize the transition states for other reactions that have similar electrostatic and geometric properties, or can even be used to carry out different functions in different reaction types[26].

Phosphoryl and sulfuryl transfer reactions are good examples of reaction classes that proceed through similar negatively charged, trigonal bipyramidal transition states. An example of more strikingly divergent reactivity is that of vanadium-dependent chloroperoxidase and the PAP2 acid phosphatases (**Fig. 2**). The acid phosphatases are inhibited by vanadate via formation of a trigonal bipyramidal species bound to the active site histidine nucleophile. The homologous chloroperoxidase has a stably bound trigonal bipyramidal vanadyl species that is used in the peroxidase reaction, and without the vanadyl species this enzyme has a low level of promiscuous phosphatase activity[26]. Thus, common active site features are used to bring about distinct catalytic outcomes.

Examples abound of modern enzymes that exhibit catalytic promiscuity and are evolutionarily related to other enzymes that proficiently catalyze that activity. These observations leave an important question unanswered, however: did the common ancestor of these enzymes exhibit promiscuity? Recently it has become possible to experimentally address these questions through the power of ancestral gene reconstruction. Using maximum-likelihood methods, ancestral DNA sequences can, in favorable circumstances, be deduced, synthesized and made to express the ancestral protein. Currently the best evidence for promiscuity in ancestral proteins comes not from an enzymatic system but from a family of steroid receptors in which the ancestral gene product was shown to have promiscuous binding activity for a steroid that did not emerge until millions of years after the ancestral receptor[28]. Given the tight link between binding and catalysis, this result has clear implications for promiscuity in ancestral enzymes. More importantly, it is now possible to move beyond sequence comparisons to functional tests of evolutionary models that directly probe the chemical changes that took place as new functions evolved[29]. Conversely, better understanding of evolutionary processes and functional development may continue to reveal basic principles of enzymatic function and may help inform strategies for the selection and design of new enzymes.

From the standpoint of understanding modern-day biological processing, recognizing that biological systems have been shaped by evolution rather than design informs our understanding of their underlying logic and organization. We can provide clear, chemical rationale for why enzymes that catalyze completely distinct reactions share sequence and structural homology (for example, **Fig. 2**), or why the ribosome is a ribozyme rather than a protein enzyme. As we strive to understand the organizing principles of more complex systems, we must continue to be informed by the precepts of molecular evolution rather than simple design criteria.

Outlook

The success of enzymology in the past half-century has been astounding. We have sufficient understanding of bioorganic chemistry and of enzyme active sites to write a reasonable chemical mechanism for nearly all enzymatic reactions (**Fig. 3**). We have a good sense

of the physical properties responsible for the enormous rate enhancements and exquisite specificities of enzymes. Nevertheless, many important challenges remain in these areas, and we hope that the successes of the past do not temper enthusiasm for continued investigation of the remaining mysteries.

Above we have highlighted some examples of the broad impact of enzymology at a level beyond simply examining individual enzymes and the chemical reactions they catalyze. Although there are other examples, we suggest that the impact of enzymology for the typical biologist and for cell biology questions has been limited. We are not certain of the reasons for this, but they may include (i) distinct training for enzymologists and biologists, such that a common language is lacking, and (ii) the richness of research questions within traditional realms of each area. And indeed there have been increasing mechanistic perspectives brought to bear on higher-order questions in biology, prompting new understanding of the capabilities and behaviors of biological systems at a level that could not be obtained simply from knowing the important players and their interacting partners. But, in relation to the modern opportunities to study function built on the enormous mass of genomic and genetic data, these approaches are vastly underrepresented in current biology research.

And what of chemical biology? We believe that this new field is at a crossroads. Currently those in this field, or who consider themselves part of this field, generally use synthetic tools from chemistry to create tools for biological investigation (**Fig. 3**). Some of these tools have been extremely powerful; others remain unproven. Thus, while interdisciplinary, chemical biology has taken a slice from each discipline rather than more fully encompassing and unifying these areas.

We believe a broader definition of chemical biology will ultimately allow broader and more incisive utilization of chemical tools and deeper insights into biology. In essence, most chemical approaches mimic and aim to improve existing molecular biology and genetic approaches, such as the ability to turn genes on and off and to screen genes and gene products for involvement in particular processes. But there are biological questions that are uniquely chemical, and these will require new focus.

Indeed, the natural language to describe reactions, individual or in combination, is in the chemical language of thermodynamics and kinetics. It is now clear, as it as been for decades[17], that biological understanding will require grasping the pathways and systems of reactions, their control and their interrelationships. Knowing the cast of characters, or even what scenes they appear in, is an essential starting point, but ultimately is not enough. Future progress will require purification, reconstitution and measurements of concentrations, equilibrium constants and rate constants—the traditional tools of the trade of the enzymologist. In this effort, current advances in single-molecule fluorescence approaches build on enzymology and may be needed to sort out the enormous complexity of biological systems. In addition, the tools of chemical biology may aid in the kinetic and thermodynamic dissection of pathways and may prove particularly powerful in the critical step of linking the behavior of reconstituted and cellular systems.

ACKNOWLEDGMENTS

We thank P. Sigala, J. Lassila, M. Good, W. Lim and G. Narlikar for their comments on this manuscript. The authors are grateful to the Damon Runyon Cancer Research Foundation (J.G.Z.) and the US National Institutes of Health (D.H.) for support.

1. Bruice, T.C. & Benkovic, S.J. *Bioorganic Mechanisms* (W.A. Benjamin, Inc., New York, 1966).
2. Kornberg, A. *Methods Enzymol.* **182**, 1–5 (1990).
3. Jencks, W.P. *Catalysis in Chemistry and Enzymology* (Dover Publications, Inc., New York, 1987).
4. Fersht, A. *Structure and Mechanism in Protein Science* (W.H. Freeman and Company, New York, 1999).
5. Bell, R.M. & Koshland, D.E. *Science* **172**, 1253–1256 (1971).
6. Voet, J.G. & Abeles, R.H. *J. Biol. Chem.* **245**, 1020–1031 (1970).
7. Fisher, H.F., Conn, E.E., Vennesland, B. & Westheimer, F.H. *J. Biol. Chem.* **202**, 687–697 (1953).
8. Kokesh, F.C. & Westheimer, F.H. *J. Am. Chem. Soc.* **93**, 7270–7274 (1971).
9. Polanyi, M. *Z. Elektrochem.* **27**, 142–150 (1921).
10. Wolfenden, R. *Annu. Rev. Biophys. Bioeng.* **5**, 271–306 (1976).
11. Pauling, L. *Chem. Eng. News* **24**, 1375–1377 (1946).
12. Page, M.I. & Jencks, W.P. *Proc. Natl. Acad. Sci. USA* **68**, 1678–1683 (1971).
13. Jencks, W.P. *Adv. Enzymol. Relat. Areas Mol. Biol.* **43**, 219–410 (1975).
14. Jencks, W.P. *Methods Enzymol.* **171**, 145–164 (1989).
15. Storm, D.R. & Koshland, D.E. *Proc. Natl. Acad. Sci. USA* **66**, 445–452 (1970).
16. Bruice, T.C., Brown, A. & Harris, D.O. *Proc. Natl. Acad. Sci. USA* **68**, 658–661 (1971).
17. Tiselius, A. in *Nobel Lectures, Chemistry 1942–1962* 195–215 (Elsevier, Amsterdam, 1964).
18. Benkovic, S.J., Valentine, A.M. & Salinas, F. *Annu. Rev. Biochem.* **70**, 181–208 (2001).
19. Kuriyan, J. & O'Donnell, M. *J. Mol. Biol.* **234**, 915–925 (1993).
20. Trahey, M. & McCormick, F. *Science* **238**, 542–545 (1987).
21. Pollard, T.D. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 451–477 (2007).
22. Rust, M.J., Markson, J.S., Lane, W.S., Fisher, D.S. & O'Shea, E.K. *Science* **318**, 809–812 (2007).
23. Orgel, L.E. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).
24. Lincoln, T.A. & Joyce, G.F. *Science* **323**, 1229–1232 (2009).
25. Narlikar, G.J. & Herschlag, D. *Annu. Rev. Biochem.* **66**, 19–59 (1997).
26. O'Brien, P.J. & Herschlag, D. *Chem. Biol.* **6**, R91–R105 (1999).
27. Khersonsky, O., Roodveldt, C. & Tawfik, D.S. *Curr. Opin. Chem. Biol.* **10**, 498–508 (2006).
28. Dean, A.M. & Thornton, J.W. *Nat. Rev. Genet.* **8**, 675–688 (2007).
29. Aharoni, A. *et al. Nat. Genet.* **37**, 73–76 (2005).
30. Pettersen, E.F. *et al. J. Comput. Chem.* **25**, 1605–1612 (2004).