Cutting the forest to see a single tree?

Antoine M van Oijen

The development of single-molecule tools has significantly impacted the way we think about biochemical processes. Watching a single protein in action allows us to observe kinetic details and rare subpopulations that are hidden in ensemble-averaging techniques. I will discuss here the pros and cons of the single-molecule approach in studying ligand binding in macromolecular systems and how these techniques can be applied to characterize the behavior of large multicomponent biochemical systems.

The ability to study the behavior of systems at the single-molecule level has had great impact on our understanding of the inner workings of biological macromolecules. Without the ensemble averaging inherent to bulk-phase biochemical assays, distributions of molecular properties can be characterized and rare subpopulations can be identified. For example, enzymes purified to homogeneity and obeying classical Michaelis-Menten behavior have been shown to display large variations in turnover rate at the single-molecule level, both over time and between molecules[1–4]. A second important benefit of the single-molecule approach as it relates to understanding mechanistic properties of enzymatic systems is the ability to observe the temporal evolution of complicated biochemical reactions. The recording of these 'molecular movies' may reveal the presence of fleeting reaction intermediates whose existence is too short-lived to be visualized in an ensemble of asynchronous reactions. The strength of this approach is nicely illustrated in the study of the stepping behavior of individual molecular motors such as myosins, kinesins, RNA polymerases and helicases[5–9].

Single-molecule tools promise to be particularly powerful in characterizing systems in which multiple binding partners interact. How is cooperativity in binding of a substrate to an enzyme regulated? How do multiple, identical proteins bind to form filaments?

Antoine M. van Oijen is at Harvard Medical School, Department of Biological Chemistry and Molecular Pharmacology, 240 Longwood Avenue, SGM204A, Boston, Massachusetts 02115, USA.
e-mail: antoine_van_oijen@hms.harvard.edu

"So in other words, we're hoping to discover what makes the nitty, gritty."

How do multiple, different proteins combine to form functional multiprotein machineries? On the one hand, single-molecule techniques enable a direct and unobscured observation of intermolecular interactions in multicomponent systems. On the other hand, there exists an inherent tension between the strategy of observing individual molecules at vanishingly low concentrations and the need for proteins to be present at sufficiently high concentrations to form complexes. I will discuss recent developments in the single-molecule field that aim toward the characterization of multicomponent systems, and I will address some key kinetic and thermodynamic considerations.

The fluctuating enzyme

A number of groups have recently visualized individual catalytic turnovers of single enzymes[2,3,10,11]. In these experiments, enzymes are immobilized on a surface and monitored using fluorescence microscopy. By using substrates that become fluorescent only after enzymatic catalysis has taken place, product generation can be observed in the presence of high substrate concentration. A vanishingly low surface density of immobilization allows enzymes to be spatially well separated and allows time series of turnovers catalyzed by a single enzyme to be recorded. The turnover rate is measured in real time by counting the number of fluorescence bursts per time unit at the enzyme's location. In college biochemistry classes we are taught to think about enzymatic kinetics from the point of view of an ensemble: how much substrate is turned over by a given amount of enzyme after a certain amount of time. English *et al.* characterized the kinetics of β-galactosidase at the single-molecule level and provided a framework for thinking about enzymatic kinetics not in the 'classical' Michaelis-Menten sense of relating enzymatic rate with substrate concentration, but from a single-molecule perspective by analyzing the waiting times between subsequent individual turnover events as a function of substrate concentration[3]. This approach allows researchers to think about enzymatic kinetics from a time perspective as opposed to a rate perspective—arguably a more intuitive way. Using enzyme cycling times instead of reaction fluxes as the observable for enzymatic activity is particularly suited for dealing with the stochastic nature of data produced by single-enzyme experiments.

Perhaps the most surprising aspect emerging from these and other single-enzyme studies is that the turnover rate of a single enzyme can fluctuate significantly over time.

In **Figure 1a**, the turnover rate of a single β-galactosidase enzyme is depicted as a function of time. The turnover rate can change by as much as a factor of ten, and the typical timescales over which these fluctuations occur have been reported to vary between milliseconds and minutes[2,3]. It has been suggested that underlying such a memory effect are small conformational changes in the enzyme. This 'breathing' results in small changes in active site geometry that in turn would modulate the substrate affinity or the reaction kinetics. Other single-molecule experiments have confirmed the existence of such small conformational changes on a variety of timescales[12].

The interconversion of an enzyme between conformational states each with different substrate-binding properties can have dramatic consequences in terms of cooperative behavior. Imagine a single-site enzyme with two initial states, one with a much higher affinity for substrate than the other. If the enzyme 'resets' to the tight-binding mode after completion of a round of catalysis (**Fig. 1b**), a situation can be achieved in which the enzymatic activity displays a sigmoidal dependence on substrate concentration (**Fig. 1c**)[13,14]. This 'dynamic' cooperativity would be apparent in a substrate concentration regime in which binding to the tight-binding state of the enzyme (k_1[S] in **Fig. 1b**) occurs on a timescale faster than equilibration between the two states (indicated by the rate constants α and β in **Fig. 1b**). At lower substrate concentrations, the two states would have equilibrated before substrate binding, thereby resulting in a lower affinity on average.

This perhaps counterintuitive behavior results in a picture in which a fluctuating, single-site enzyme uses a substrate-substrate interaction through time to achieve cooperativity, in stark contrast to the classical notion of allosteric interactions between multiple sites on a protein (with hemoglobin as the most well-known example). Potentially, this mechanism could be of importance in the regulation of the kinetics of complex networks of protein-substrate interactions. Future studies of enzyme kinetics, using both single-molecule and bulk-phase methods, are needed to further explore such dynamic cooperativity.

Cutting the trees

In a way, the single-molecule approach seems to represent a double-edged sword: observation of the kinetics of a biochemical reaction with single-molecule sensitivity and resolution can most often only be done at vanishingly low concentrations of one of the reactants. In experiments in which fluorescence is used as a probe, the concentration of fluorescently labeled species typically cannot be much higher than 1 nM, which is much lower than the typical affinity between two biological binding partners. In other single-molecule approaches, fluorescence is not used; instead, a mechanical force is exerted on a molecule and enzymatic activity is measured as a mechanical change in the system (for a review of different single-molecule approaches, see ref. 15). In these experiments, it is the mechanically manipulated molecule that is often present at extremely low concentrations (often as low as one copy of the molecule in the entire experimental volume).

The reason that these single-molecule experiments work and give a bona fide representation of the kinetic and thermodynamic properties of the system is that the concentration of only the component that gives rise to the experimental observable needs to be low. The concentration of the other binding partner, for example, a substrate for an enzymatic reaction, can be varied over a wide range to provide experimental access to systems with many different binding and kinetic behaviors. As long as only one of all binding partners in a complex-formation process is present at a concentration much lower than the binding affinity between the partners, the binding rates and fractional occupancy are independent of the actual molarity of the low-concentration species. For example, the fractional occupancy of a binding site B with a ligand L is expressed as

$$c \equiv \frac{[B \cdot L]}{[B_{\mathrm{tot}}]} = \frac{[L]}{K_{\mathrm{d}}+[L]}$$

where K_{d} is the equilibrium dissociation constant of the interaction and $[L]$ is the concentration of unbound ligand (equal to $[L_{\mathrm{tot}}]$ minus $[B{\cdot}L]$). Under the condition $[L_{\mathrm{tot}}] >> [B_{\mathrm{tot}}]$, $[L]$ can be approximated by $[L_{\mathrm{tot}}]$ and the fractional occupancy becomes independent of the concentration of binding sites.

Care should be taken, however, in directly comparing results of single-molecule binding studies with those obtained using bulk-phase assays. Binding studies in the bulk phase are often done at binding-site concentrations that are not insignificant compared to the equilibrium dissociation constant. As a result, a molecule with binding site B present in extremely low concentrations in a single-molecule experiment can reach a much higher degree of occupancy than each of the single molecules in a bulk-phase experiment at identical ligand

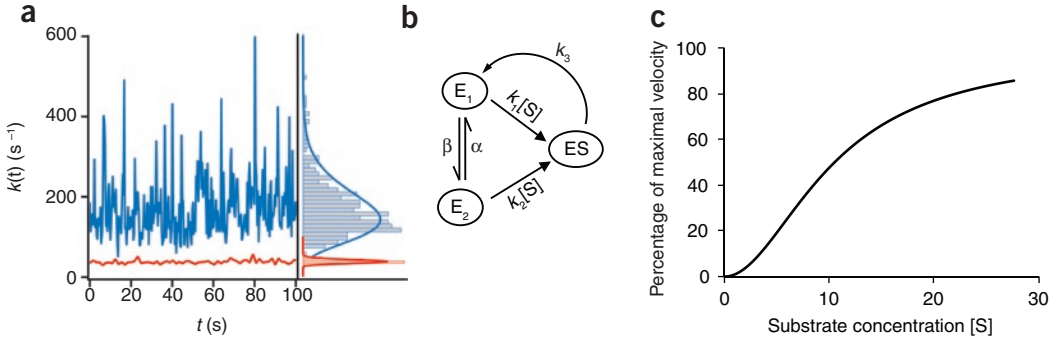

Figure 1 Fluctuating turnover rates and binding affinities of enzymes. (**a**) Experimental trace of the turnover rate of a single β-galactosidase enzyme as a function of time. Blue and red traces each correspond to a single enzyme. The histograms represent the distribution of rates for each enzyme integrated over time. The difference between the means of the two distributions indicates the large variation in turnover rate between two enzymes. The fluctuations in the blue trace and the large width of the blue histogram indicate large fluctuations of the turnover rate of an individual enzyme over time. (**b**) Enzymatic reaction scheme with enzyme E in one of two conformational states, E_1 and E_2, each with different affinities for substrate S. Rate constants α and β indicate interconversion kinetics between E_1 and E_2. After binding of substrate S, enzyme-substrate complex ES is formed, turnover takes place and the enzyme reverts to the E_1 state. (**c**) The steady state velocity of the enzymatic reaction depicted in **b**, with $k_1 = 100$, $k_2 = 0.01$, $\alpha = 0$ and $\beta = 1$. The sigmoidal shape indicates cooperative behavior. **a** is from ref. 3; **b** and **c** are adapted from ref. 14.

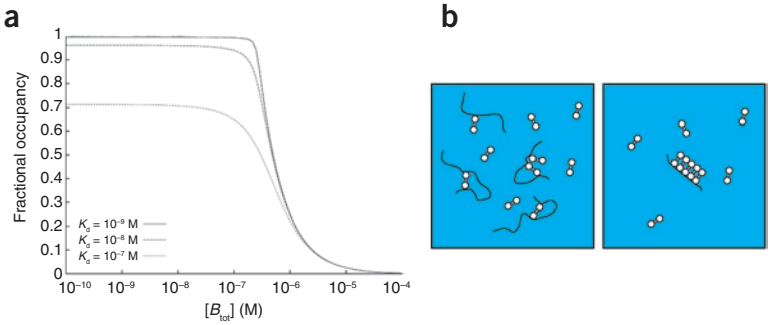

Figure 2 Fractional binding in single-molecule and bulk-phase experiments. (**a**) Fractional occupancy of binding sites as a function of concentration of binding sites [*B*] for three different values of K_d. The ligand concentration [*L*] is held constant at 2.5×10^{-7} M. At a fixed ligand concentration, each of the binding sites will be maximally occupied when [*B*] decreases to the single-molecule limit. In this regime, the fractional occupancy is not sensitive anymore to changes in the ligand concentration or K_d. Figure adapted from ref. 20. (**b**) Fractional occupancy of DNA by H-NS increases as DNA concentration decreases at constant H-NS concentration. Where a bulk-phase biochemical assay may report on a low occupancy at high DNA concentration (left), a single-molecule experiment with as little as one DNA molecule may give rise to significantly altered binding properties (right). In the case of H-NS, a high occupancy could inhibit the formation of intrastrand loops in the DNA substrate.

concentrations (**Fig. 2a**). An example of a system that is reported to display starkly different behavior in single-molecule and bulk-phase experiments is the binding of the *Escherichia coli* histone-like nucleoid structuring protein (H-NS) to DNA. H-NS plays a role in both structuring DNA and regulating transcription in Gram-negative bacteria[16]. The protein exists as a dimer, with each subunit containing a DNA-binding domain. After initial binding to the DNA of one of the DNA-binding domains, the protein mediates loop formation by an interaction of the second DNA-binding domain to a distal site on the duplex. Interestingly, single-molecule magnetic tweezer experiments that measured the length of individual DNA molecules as a function of H-NS concentration resulted in the observation that H-NS induces a rigidification of the DNA molecule[17]. This observation was in contrast to previous studies that demonstrated a compaction of DNA upon association with H-NS[18]. In a polemic series of letters following the publication of the original single-molecule paper[17], it was argued that a large excess of protein over DNA led to a rapid saturation of the available binding sites on the DNA and caused the DNA to stiffen before intrastrand contacts could be made and loops formed (**Fig. 2b**)[19,20].

In general, there can be many reasons why single-molecule and ensemble biochemical experiments may give rise to conflicting answers. It is important to be continuously aware of the large concentration differences between *in vitro* single-molecule experiments and bulk-phase assays. Experimental conditions of binding assays need to fulfill both the criteria of a sufficiently high ligand concentration and an appropriate stoichiometry of binding partners. Whereas the first criterion is easily met in single-molecule experiments, the latter is often difficult to achieve because of a need for a 'single-molecule' concentration of at least one of the components. Before interactions between proteins are characterized at the single-molecule level, altered concentration requirements should be tested at the ensemble level for their ability to support appropriate reconstitution of the biological activity under investigation. More generally, single-molecule experiments can often only be designed properly and executed successfully if a significant amount of biochemical information is already known. In particular for larger, multiprotein complexes, a good understanding of affinities and rate constants needs to exist before the right single-molecule conditions can be chosen. Of the many possible observables in a single-molecule experiment, at least one should be chosen such that a direct comparison with bulk-phase data is possible. Whereas single-molecule experiments can give great insight into the distribution of a particular molecular property, the mean of that distribution should be consistent with data from conventional biochemistry.

In some cases, however, a direct comparison can be misleading, and average values obtained from single-molecule experiments may provide a more accurate readout of the activity of a system than results from bulk-phase experiments. For instance, measurements of processivity of nucleic acid enzymes, such as DNA polymerases, are challenging to obtain at the ensemble level. The use of ensemble-averaging techniques to measure processivity values is often complicated by the occurrence of multiple consecutive events of binding, synthesis and dissociation on one and the same DNA molecule—a situation that is difficult to distinguish from a single highly processive event. The use of 'trap' DNA to bind the enzyme the moment it dissociates from its original template is a standard trick in the repertoire of bulk-phase biochemistry to ensure that each enzyme only mediates one cycle of binding, enzymatic activity and unbinding. A shortcoming of this method is that even at micromolar trap concentrations, the average distance between the enzyme-substrate complex and a trap molecule is still many tens of nanometers. After dissociation from its substrate, diffusion can bring the enzyme back to its original substrate where it can undergo another burst of enzymatic activity, thereby giving rise to inflated processivity readouts. The real-time observation of the interaction between enzyme and DNA at the single-molecule level provides a much more direct readout of enzymatic properties and gives a significantly less convoluted view on parameters such as binding rate, turnover rate, processivity and dissociation rate[21].

Assembly and functioning of large multiprotein machineries

The field of single-molecule biophysics has seen tremendous advances in the study of individual enzymes and the understanding of their catalytic mechanisms. In a biological context, most of these enzymes function in concert with other enzymes in multiprotein complexes, so an important future direction will be the use of single-molecule techniques to unravel the orchestration of large macromolecular assemblies. As indicated above, studying the interaction between a protein and its substrate at the single-molecule level is readily achievable by decreasing the concentration of one of the two to a level amenable to single-molecule detection techniques. Upon introducing more binding partners to form a larger complex, however, the concentration of every additional component needs to be at least comparable to the K_d between that component and the rest of the complex to ensure proper association of all the binding partners.

In practice, the members of a multiprotein complex all need to be present in sufficiently high concentration to allow complex formation at reasonable timescales. Some experiments on larger, multi-enzyme complexes have already been successfully done by maintaining protein concentrations sufficiently high to mediate complex formation and by keeping the substrate concentration low. For example, the activity of prokaryotic DNA replication machineries has been characterized by measuring the length of individual DNA substrates as the replication complex, containing a helicase,

primase, DNA polymerases and single-strand binding proteins, duplicates the parental strand[21,22]. The different elastic properties of double- and single-stranded DNA, and the looping of the DNA by the replication machinery, make it possible to derive enzymatic activity by monitoring small changes in the length of individual DNA molecules[23]. The dynamic properties of replication complexes containing up to a dozen protein components have been studied using this method.

The next logical step in the single-molecule characterization of large multiprotein complexes is the use of fluorescence techniques to understand the microscopic details of how the proteins interact in larger complexes. For example, single-molecule fluorescence resonance energy transfer (FRET) can be used to probe subtle distance changes between components[24]. However, the requirement for high protein concentration is seemingly incompatible with single-molecule fluorescence detection. When imaging the fluorescence of labeled protein, it is only possible to detect individual proteins when the concentration of labeled species is around a few nanomolars or lower—a concentration lower than the K_d of most protein-protein interactions. This upper limit represents the concentration at which one molecule is maximally present in one optical detection volume. The development of physical techniques that further reduce the observation volume in fluorescence microscopy may lead to a loosening of this requirement[25,26]. An exciting direction is the development of optical schemes that reduce the physical size of the excitation volume created by a laser focus. Not only will this method increase the spatial resolution in fluorescence microscopy, with obvious impact in many fields related to imaging, but it will also significantly lower the number of emitting fluorophores contributing to background signal and allow single-molecule experiments at higher concentrations[26].

Another strategy is the reduction of the physical volume that contains the reactants. Early work in this direction revolved around the fabrication of microfluidic reaction chambers whose dimensions are significantly smaller than the resolution limit of optical microscopes[25]. More recently, researchers have been able to capture fluorescently labeled binding partners in small (100–200 nm), surface-tethered lipid vesicles and study their interactions through single-molecule FRET measurements[27,28]. The volume of a 100-nm vesicle is two orders of magnitude smaller than that of a diffraction-limited, confocal excitation focus and thus allows proteins to be visualized at the single-molecule level at much higher concentrations. These approaches are opening the way to studying interactions between molecules that have K_ds much weaker than previously thought to be accessible by single-molecule techniques. Extrapolating the strategy of increasing the local concentration of binding partners while maintaining a low number of fluorescently labeled species in the observation volume, one can envision methods where binding partners with a weak interaction are covalently coupled through a tether. Physically constraining the distance between two proteins to a few nanometers will enable the single-molecule characterization of weak interactions with K_ds in the millimolar regime.

In summary, single-molecule techniques have allowed researchers to approach the concept of multi-site binding and assembly from an entirely different perspective. Single-molecule enzymology demonstrated dynamic effects that were hidden from view entirely by ensemble-averaging techniques. The flipside of the coin is the requirement to work at concentrations much lower than those used in bulk-phase biochemical experiments and perhaps also much lower than physiological concentrations. Whereas for simple binding reactions with a limited number of binding partners this is not necessarily a problem, with more complicated systems there exists an inherent tension between the extreme experimental dilutions on the one hand and the limited affinities between proteins on the other. The pace with which new physical tools and novel labeling strategies find their way into single-molecule biochemistry will hopefully lead to a situation where single-molecule approaches can be routinely applied to large complexes in physiologically relevant settings.

ACKNOWLEDGMENTS

The author acknowledges funding from the US National Institutes of Health, the US National Science Foundation and the Searle Scholars Program. The author thanks J. Loparo for a critical reading of the manuscript.

1. Rissin, D.M., Gorris, H.H. & Walt, D.R. *J. Am. Chem. Soc.* **130**, 5349–5353 (2008).
2. Flomenbom, O. *et al. Proc. Natl. Acad. Sci. USA* **102**, 2368–2372 (2005).
3. English, B.P. *et al. Nat. Chem. Biol.* **2**, 87–94 (2006).
4. van Oijen, A.M. *et al. Science* **301**, 1235–1239 (2003).
5. Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R. & Block, S.M. *Nature* **438**, 460–465 (2005).
6. Dumont, S. *et al. Nature* **439**, 105–108 (2006).
7. Myong, S., Rasnik, I., Joo, C., Lohman, T.M. & Ha, T. *Nature* **437**, 1321–1325 (2005).
8. Yildiz, A. *et al. Science* **300**, 2061–2065 (2003).
9. Yildiz, A., Tomishige, M., Vale, R.D. & Selvin, P.R. *Science* **303**, 676–678 (2004).
10. Edman, L. & Rigler, R. *Proc. Natl. Acad. Sci. USA* **97**, 8266–8271 (2000).
11. Lu, H.P., Xun, L. & Xie, X.S. *Science* **282**, 1877–1882 (1998).
12. Yang, H. *et al. Science* **302**, 262–266 (2003).
13. Cornish-Bowden, A. & Cardenas, M.L. *J. Theor. Biol.* **124**, 1–23 (1987).
14. Qian, H. *Biophys J.* published online, doi:10.1529/biophysj.108.131771 (25 April 2008).
15. Cornish, P.V. & Ha, T. *ACS Chem. Biol.* **2**, 53–61 (2007).
16. Rimsky, S. *Curr. Opin. Microbiol.* **7**, 109–114 (2004).
17. Amit, R., Oppenheim, A.B. & Stavans, J. *Biophys. J.* **84**, 2467–2473 (2003).
18. Tupper, A.E. *et al. EMBO J.* **13**, 258–268 (1994).
19. Amit, R., Oppenheim, A.B. & Stavans, J. *Biophys. J.* **87**, 1392–1393 (2004).
20. Dame, R.T. & Wuite, G.J. *Biophys. J.* **85**, 4146–4148 (2003).
21. Tanner, N.A. *et al. Nat. Struct. Mol. Biol.* **15**, 170–176 (2008).
22. Lee, J.-B. *et al. Nature* **439**, 621–624 (2006).
23. van Oijen, A.M. *Biopolymers* **85**, 144–153 (2007).
24. Ha, T. *Methods* **25**, 78–86 (2001).
25. Levene, M.J. *et al. Science* **299**, 682–686 (2003).
26. Klar, T.A., Jakobs, S., Dyba, M., Egner, A. & Hell, S.W. *Proc. Natl. Acad. Sci. USA* **97**, 8206–8210 (2000).
27. Benitez, J.J. *et al. J. Am. Chem. Soc.* **130**, 2446–2447 (2008).
28. Cisse, I., Okumus, B., Joo, C. & Ha, T. *Proc. Natl. Acad. Sci. USA* **104**, 12646–12650 (2007).