Primer

Simulating cell biology

Steven S. Andrews[1] and Adam P. Arkin[1,2]

Science is an iterative process of experiments and hypotheses. Experiments produce surprising results; hypotheses are created to explain the results; new experiments are designed to test the hypotheses, of which some agree, some fail without yielding useful information and some produce more surprises; and the cycle continues. As a field matures, knowledge grows and the hypotheses become more elaborate, eventually exceeding the limits of what a scientist can mentally grasp. This is where computational modeling becomes necessary, and where cell biology is today.

Modeling serves the same purposes as scientific cartoons or calculations on the backs of envelopes, but is much more precise. A model can definitively show if an hypothesis can explain a set of data, make experimental predictions and help identify system aspects that are poorly understood. After many iterations of experiments and theory, models are often sufficiently supported by evidence that they represent the current understanding of a system, against which new results are compared. This primer focuses on simulations of biochemical reaction networks, which is a core component of most cell biological models. We leave aside the related arts of model building, model analysis such as sensitivity analysis, and model/data comparison.

We explore a range of simulation methods that vary in their level of physical approximation and abstraction. For concreteness, each is presented for the same generic elementary reversible chemical reaction:

$$A + B \underset{k_r}{\overset{k_f}{\rightleftarrows}} C \qquad (1)$$

These molecules might be proteins, small molecules, DNA, RNA or other species. The molecular concentrations, denoted below as a, b and c, are variables that change over time, whereas the system volume, the initial concentrations and the reaction rate constants k_f and k_r are physical parameters that are specified by the modeler and kept constant throughout the simulation. When these parameters have not been directly measured and cannot be estimated from indirect data, model behavior is often explored as a function of them. Outputs of the simulation methods discussed below, applied to a simple chemical oscillator, are compared in Figure 1. As this is a primer rather than a review, we only list a few of the many excellent references.

ODE models

One of the most approximate physical models of biochemical networks, the kinetic ordinary differential equation (ODE), assumes that molecular concentrations are continuous (it ignores the discrete nature of molecules), that reactions occur in a homogeneous, well-stirred volume and that these reactions occur in a deterministic manner. This is by far the most common form of biological model and can represent both the transient dynamics and the long-term steady-state behavior of a system if the above approximations hold.

ODE models of biochemical networks have been successfully applied to diverse systems. Metabolic networks are often investigated under constant conditions to identify the steady-state rates of metabolite flux and cellular growth [1]. Switch-like memories, involving bistability and hysteresis, are observed in cell cycle models [2]: the division control machinery is switched from one stable set of chemical concentrations during interphase to another during early M phase by a high concentration of cyclin; after division, the cyclin concentration crosses a low threshold to switch the system back to interphase. Oscillations are found, not surprisingly, in circadian rhythm models [3]. Finally, ODEs can display deterministic chaos [4], which has been found in simulated reaction networks but not live cells, perhaps because of evolutionary selection against it.

In an ODE model, a reaction network is expressed as a set of differential equations with one equation per chemical and with terms that represent the reactions. Using A+B↔C, and assuming mass action kinetics, the differential equation for C is:

$$\frac{dc}{dt} = k_f ab - k_r c \qquad (2)$$

Equations for a and b are analogous: both are the negative of equation 2. The time dependence of the chemical concentrations, and the long-term states, can be solved analytically for very simple cases and numerically [5] for more typical networks. These numerical methods, along with those described below, span a wide range of sophistication to control different sources of numerical and approximation error. Simple ones are often adequate for initial work, but they run slowly and are prone to unstable behavior, while complex ones can be research programs in themselves.

Spatial models

It is becoming increasingly clear that, even in bacteria, there is an exquisite spatial organization of cellular components, and that the dynamic localization and formation of proteins and cellular superstructures plays a key role in cellular processes ranging from cell shape, to cell cycle, to signaling. This organization can only be modeled by accounting for space as well as reactions.

Spatial models can represent the same behaviors listed above, plus several new ones. Chemical waves are seen in neuroblastoma cells [6]: a calcium burst starts in the center of a neurite, which starts a wave that propagates outwards towards the soma and

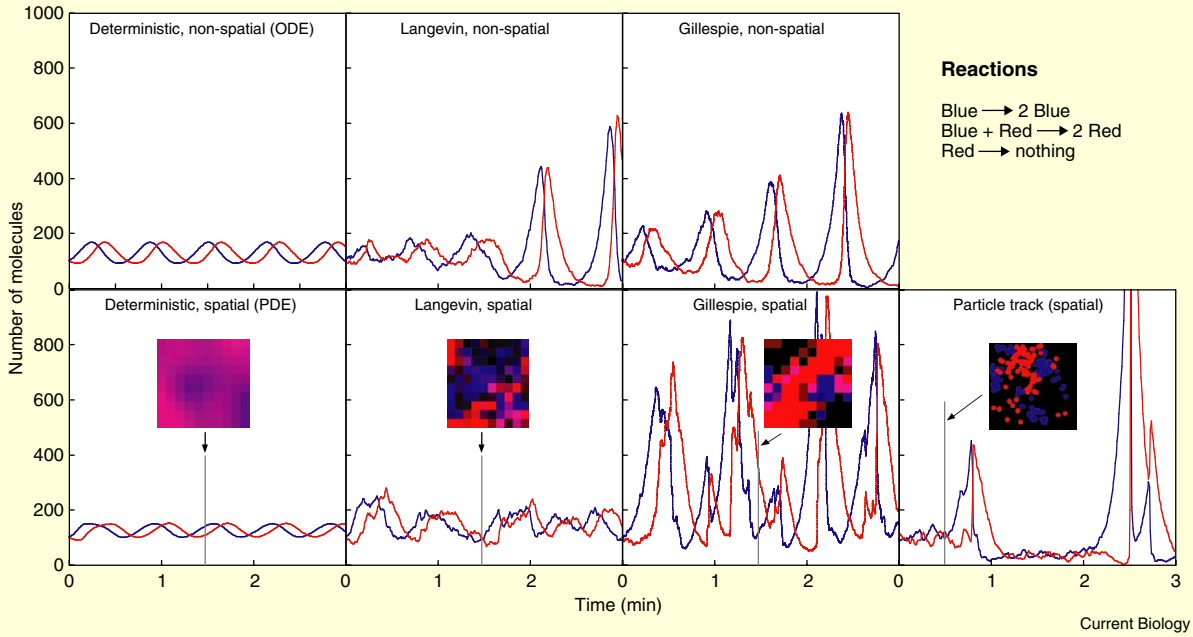

Figure 1. Simulation results for a simple chemical oscillator using different simulation methods.

The Lotka-Volterra system is shown, which shares key features with cellular oscillators, such as circadian rhythms. Insets show the spatial distributions of molecules at the indicated times. In the top panels, note that stochasticity allows the system to drift to large amplitude oscillations and that the Langevin and Gillespie methods yield similar results. In the bottom panels, all of which were started with nearly homogeneous initial states, differences arise from the approximations: the PDE simulation has predictable oscillations due to the minimal stochasticity (which is only in the initial state); the Gillespie simulation has larger peaks than the Langevin one because it only allows integer numbers of molecules in each bin; and the particle tracking simulation shows larger and fewer bursts than does the Gillespie simulation because it accurately treats diffusion at all length scales (this difference was reduced with a spatial Gillespie simulation that used smaller subvolumes). Parameters: rate constants are 10 min^{-1}, 8000 $nm^3molec^{-1}min^{-1}$, and 10 min^{-1}, for the respective reactions shown in the top-right corner, systems start with 100 of each blue and red molecules, their diffusion coefficients are 100 nm^2min^{-1}, the volume is 100 nm high and wide by 10 nm deep, and the first three spatial simulations divide this volume into cubic subvolumes that are 10 nm on a side. Simulations were performed with the Smoldyn program, which can be downloaded from http://genomics.lbl.gov/~sandrews/software.html.

growth cones. Spatial oscillations are used by *Escherichia coli* to center the cell division site [7]: oscillation of MinC protein back and forth in the cell, arising from dynamics of the MinD and MinE proteins, causes its time-averaged concentration to be high at the cell poles and low at the center; this directs the division apparatus to the center. Spontaneous pattern formation [8] arises in the development of an embryo from an egg: a reaction network of morphogens is sufficiently unstable in the initial symmetric system that random perturbations trigger pattern formation, which is used to position the head, tail, front, and back of the embryo. Each of these models showed how the known biochemistry for the system could lead to the observed dynamics.

Spatial phenomena arise when the timescales of diffusion are somewhat slower than those for reactions. Diffusion timescales are about L^2/D, where D is a molecule's diffusion coefficient and L is a characteristic length such as the cell length or the distance between concentration waves. Reaction timescales, using A+B↔C, are $(a+b)/(k_f ab)$ for the forward reaction and k_r^{-1} for the reverse reaction.

If the system consists of several compartments, for example nucleus and cytoplasm, and there is rapid mixing within compartments, then simulation is easiest with a compartmental model. The dynamics are treated with the same ODEs as for non-spatial models, but now with a set of equations for each compartment and also terms for transport between compartments.

To model continuous space, concentrations become functions of the position. For A+B↔C, the concentrations are $a(\mathbf{x})$, $b(\mathbf{x})$, and $c(\mathbf{x})$, where $\mathbf{x}$ is a position vector. The time dependence of C is given with the partial differential equation (PDE)

$$\frac{\partial c(\mathbf{x})}{\partial t} = k_f a(\mathbf{x}) b(\mathbf{x}) - k_r c(\mathbf{x}) + D_c \nabla^2 c(\mathbf{x}) \quad (3)$$

The first terms represent the formation and dissociation of C, now at position $\mathbf{x}$, and the last term accounts for diffusion of C into and away from $\mathbf{x}$, where D_C is the diffusion coefficient for C. Numerical integrations [8,9] treat the space by partitioning it into subvolumes. They are particularly challenging for spatial systems because instabilities arise when the time step exceeds about $\Delta x^2/D$, where Δx is the partition width and D is the diffusion coefficient of the fastest diffusing species. Thus, high spatial resolution with low prediction error requires both small subvolumes and short time steps, which make these simulations computationally intensive.

Stochastic models

ODE and PDE models are approximate because they treat molecules with continuously variable concentrations rather than the discrete entities that they actually are [10]. In reality, reactions occur as a rapid succession of separate elementary events, the exact timing of which is effectively random because of the Brownian motion of the reactants. This stochasticity is most important when the reaction network is poised near a threshold. For example, biochemical oscillators often have a sharp transition between concentrations that produce oscillations and those that do not. Just outside an oscillating regime, relatively minor stochasticity can trigger and maintain regular oscillations, called stochastic resonance, which has been found in calcium oscillations [11] and circadian clocks [12].

Some genetic switches have evolved to be sensitive enough for naturally stochastic gene expression to produce phenotypically diverse populations, which in some cases are more evolutionarily fit than their deterministic counterparts. These include the lysis–lysogeny switch in λ phage [13] and pili phase variation in parasitic bacteria. In some cases, stochasticity simply adds noise to otherwise deterministic dynamics, while in others it can fundamentally alter the types of dynamics that are possible [14].

For most systems, if there are about n molecules of some species, this value will fluctuate with standard deviation of about $n^{1/2}$. These fluctuations are typically negligible for millions of molecules in a cell (as in the case of metabolites), significant for thousands of molecules (as with signaling proteins), and very important for tens of molecules (for example mRNA). Naturally sporadic DNA transcription causes there to be few mRNA molecules in a cell of a given type; the consequent stochasticity is amplified during translation and sometimes again during downstream gene regulation. This stochastic gene expression is likely to be a primary cause of endogenous non-genetic variation between cells.

Stochastic simulation theory typically starts with the chemical master equation [10], which tracks the probabilities of all possible system states over time. Even for trivial reaction networks, it becomes so unwieldy that modelers generally choose a Monte Carlo method, in which the simulation makes random choices as it progresses, thus producing a single stochastic time course. A different time course is produced on every run, so stochastic simulations are interpreted by analyzing statistics on the results, as one does with experiments.

Rather than presenting the defining equations, we show two stochastic simulation algorithms because they are more instructive in this case. The Gillespie algorithm [15] is exact, meaning that statistics collected from simulations have been proven to be identical to those calculated from the master equation. For A+B↔C, the algorithm updates the numbers of molecules — given by A, B, and C, as they are numbers of molecules rather than concentrations — and the simulation time according to:

(1) $s = k_f AB/V + k_r C$

(2) Δt = {exponentially distributed random number with mean $1/s$}

(3) μ = {'f' with probability $k_f AB/(Vs)$, and otherwise 'r'}

(4) increase t by Δt and {decrement A and B and increment C if μ ='f'}, or {increment A and B and decrement C if μ = 'r'}

(5) repeat from step (1)

(Minor changes are required for reactions with identical reactants.) In a faster but more approximate method, random noise is added to the deterministic ODEs to yield a form of the chemical Langevin equation [16]. A simple implementation for species C is

$$\Delta C = \left(\frac{k_f AB}{V} - k_r C\right)\Delta t + \left[N_f(t)\sqrt{\frac{k_f AB}{V}} - N_r(t)\sqrt{k_r C}\right]\sqrt{\Delta t} \quad (4)$$

In the noise term, $N_f(t)$ and $N_r(t)$ are normally distributed random numbers (Gaussian with mean 0 and standard deviation 1) that are statistically independent of each other and of all $N_f(t)$ and $N_r(t)$ with other t values. This is fairly accurate if molecule counts are not too low (below ~100) and the system is not too close to a critical point in its dynamics.

Spatial-stochastic models

Even more detailed simulations account for both space and individual molecules, which can be done with either of two general schemes. The spatial Langevin and spatial Gillespie methods combine techniques listed above by running the stochastic algorithms in each of many subvolumes [17]. Diffusion is simulated by defining new 'reactions' in which molecules move between neighboring regions with rate constants $D/\Delta x^2$. These are refinements of algorithms described previously, so models can be developed incrementally, which simplifies both model parameterization and interpretation of results.

In contrast, the particle tracking scheme [18] approaches a similar level of simulation detail from the other direction, by being an approximation of a molecular picture. Simulated molecules are represented as point-like particles with continuously variable positions. Diffusion is simulated by displacing molecules by small random amounts at each time step; pairs of molecules react when they approach each other to within a 'binding radius' or when ligands diffuse into a surface that is covered with cognate receptors.

All of these methods can be valuable for systems that require spatial resolution and accurate stochastics, such as cell signaling, although their predictions can vary significantly (Figure 1). Particle tracking is typically the most accurate of these methods because it addresses the correlations that occur between reactions: ligands bind to clustered receptors repeatedly and in localized patches to create intermittent bursts of signaling activity [19], and the dissociation of a C molecule in the A+B↔C reaction leads to products that often recombine soon after their

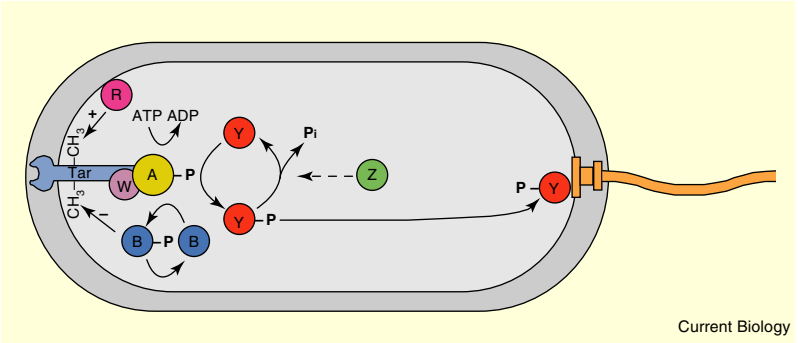

Figure 2. The *E. coli* chemotaxis signaling network, which has revealed many insights into cell biology through modeling research.

An increase of the repellent concentration promotes cell tumbling by: repellent molecules bind to receptors, they transmit a conformational shift through CheW to CheA, the activity of CheA increases, active CheA is phosphorylated, the phosphate is transferred to CheY, CheYp diffuses and binds to a motor (there are usually several flagella), the motor is biased towards clockwise rotation, and the previously bundled flagella become disordered. Dephosphorylation of CheYp by CheZ resets the system. Meanwhile, the system adapts to the repellent so that it can sense the gradient even as the concentration decreases. This happens by phosphorylation of CheB by CheAp and demethylation of receptors by CheBp, which decreases their activity; these methyls are replaced by CheR.

formation [18]. Particle tracking is also useful for simulating diffusion among inert macromolecules and convoluted membranes, which can affect biochemical signal transmission [20,21].

Mechanical models

A tremendous diversity of mechanical models have been designed, using a comparable diversity of methods. Topics that have been modeled include: actin and other polymers used in cell motility; polymers that form bacterial or eukaryotic cytoskeletons; interactions between extracellular fluid flow and cell chemotaxis; the dynamics of motor proteins; and many others. These mechanical models introduce additional variables to the chemical concentrations and chemical positions described previously, such as deformations and motions of polymers and membranes.

In many cases, mechanical processes may be simulated using technologies similar to those above, including ODEs, PDEs and stochastic methods, although they often require specialized approaches. The coupling of these processes to chemical ones can be subtle. As an example, a monomer might add to a polymer with a chemical reaction; if the polymer is in contact with a membrane, this will distort the membrane, which might then affect the growth or shrinkage rates of neighboring polymers.

Case study: chemotaxis

The *E. coli* chemotaxis signaling network (Figure 2) has been uniquely well modeled, so it is informative to trace its history and to consider a few modeling results. Briefly, *E. coli* cells are observed to swim by alternating straight runs and randomizing tumbles. Bacteria avoid a nearby repellent source by tumbling more often when they sense an increasing repellent concentration and less often when the concentration decreases. The signaling network that drives this behavior was largely figured out from experiments in the 1970s and 1980s, but was not quantitatively investigated as a system until a simulation for the initial response portion of the network was published in 1993 [22]. This model was based on ODEs, with a stochastic component for the motor complex. Some model parameters had been experimentally measured, some were found indirectly from other experiments, and others were adjusted by hand until simulation results were similar to experimental ones. The model accurately represented the phenotypes of 33 out of 41 chemotaxis mutants, but could not capture the observed high sensitivity. This landmark achievement was a typical first model: it was based on a highly simplified network, it used simple methods, and the results agreed qualitatively, but not quantitatively, with experimental ones.

Later models filled in various gaps, including the adaptation portion of the network and reactions in the receptor cluster. With the many unknown parameters, which was especially true with the newer models, parameterization became a major question. Some models optimized parameters using automated search methods, others showed that natural variability in protein concentrations can predict the phenotypic diversity found in clonal populations [23], and yet others showed that chemotaxis adaptation is remarkably insensitive to parameter values [24]. Stochastic, spatial, and particle tracking simulations have been developed for the chemotaxis network: a stochastic one explained the large temporal variations that are observed in single cells [25], a spatial one showed that the intracellular CheZ distribution affects transmitted signals [26], and a particle tracking one demonstrated the importance of macromolecular crowding [20]. Several disagreements between models and experiments have been investigated with specialized models that focused on portions of the network. In particular, receptor cluster and motor complex models showed that the high sensitivity of the network is likely to arise from allostery [27].

While all of these results were found specifically for the chemotaxis signaling network, they are likely to apply much more broadly. Thus, these models are revealing how cell biology works at a much deeper level than would have been possible from experiments alone.

Future directions

As with the experimental sciences, the cell simulation field is largely driven by the development of new techniques, and of new tools that implement those techniques. Many techniques under development

are working to increase simulation accuracy while reducing the computational burden. A central challenge concerns simulation of interacting processes that occur on different time scales; the fast scale imposes a short simulation time step, but that makes the simulation too slow to observe the slow scale. This is being met with new algorithms and hybrid simulations that treat space and stochasticity only as required. Also, given the size and the possible nonlinearity and non-determinism represented by biological models, tools for analysis of models, such as those that provide parametric sensitivity analysis, and for comparing models to data for parameterization and (in)validation are both profoundly needed and in a relatively primitive state.

It is easy to describe the ideal simulation tool: it should be able to simulate reactions and diffusion as accurately as needed, account for all relevant mechanical processes, help with model parameterization, validate and discriminate between models using data, and be easy to use. Many modeling tools are aiming towards this goal but it remains elusive, in part because of the extraordinary speed with which improved analysis methods and cellular measurements are being developed.

References

1. Covert, M.W., Schilling, C.H., Famili, I., Edwards, J.S., Goryanin, I.I., Selkov, E., and Palsson, B.O. (2001). Metabolic modeling of microbial strains in silico. Trends Biochem. Sci. *26*, 179–186.
2. Tyson, J.J. (1991). Modeling the cell division cycle: cdc2 and cyclin interactions. Proc. Natl. Acad. Sci. USA *88*, 7328–7332.
3. Goldbeter, A. (2002). Computational approaches to cellular rhythms. Nature *420*, 238–245.
4. Fox, J.J., and Hill, C.C. (2001). From topology to dynamics in biochemical networks. Chaos *11*, 809–815.
5. Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1988). Numerical Recipies in C. The Art of Scientific Computing (Cambridge, UK: Cambridge University Press).
6. Fink, C.C., Slepchenko, B., Moraru, I.I., Watras, J., Schaff, J.C., and Loew, L.M. (2000). An image-based model of calcium waves in differentiated neuroblastoma cells. Biophys. J. *79*, 163–183.
7. Huang, K.C., Meir, Y., and Wingreen, N.S. (2003). Dynamic structures in *Escherichia* coli: spontaneous formation of MinE rings and MinD polar zones. Proc. Natl. Acad. Sci. USA *100*, 12724–12728.
8. Miura, T., and Maini, P.K. (2004). Periodic pattern formation in reaction-diffusion systems: An introduction for numerical simulations. Anat. Sci. Int. *79*, 112–123.
9. Schwartz, P., Adalsteinsson, D., Colella, P., Arkin, A.P., and Onsum, M. (2005). Numerical computation of diffusion on a surface. Proc. Natl. Acad. Sci. USA *102*, 11151–11156.
10. Rao, C.V., Wolf, D.M., and Arkin, A.P. (2002). Control, exploitation, and tolerance of intracellular noise. Nature *420*, 231–237.
11. Li, H., Hou, Z., and Xin, H. (2005). Internal noise stochastic resonance for intracellular calcium oscillations in a cell system. Phys. Rev. E *71*, 061916.
12. Vilar, J.M.G., Kueh, H.Y., Barkai, N., and Leibler, S. (2002). Mechanisms of noise-resistance in genetic oscillators. Proc. Natl. Acad. Sci. USA *99*, 5988–5992.
13. Arkin, A., Ross, J., and McAdams, H.H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage l-infected *Escherichia coli* cells. Genetics *149*, 1633–1648.
14. Samoilov, M., Plyasunov, S., and Arkin, A.P. (2005). Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. Proc. Natl. Acad. Sci. USA *102*, 2310–2315.
15. Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. *81*, 2340–2361.
16. Gillespie, D.T. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. *115*, 1716–1733.
17. Elf, J., and Ehrenberg, M. (2004). Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. Syst. Biol. *1*, 230–236.
18. Andrews, S.S., and Bray, D. (2004). Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. Phys. Biol. *1*, 137–151.
19. Andrews, S.S. (2005). Serial rebinding of ligands to clustered receptors as exemplified by bacterial chemotaxis. Phys. Biol. *2*, 111–122.
20. Lipkow, K., Andrews, S.S., and Bray, D. (2004). Simulated diffusion of CheYp through the cytoplasm of *E. coli*. J. Bact. *187*, 45–53.
21. Coggan, J.S., Bartol, T.M., Esquenazi, E., Stiles, J.R., Lamont, S., Martone, M.E., Berg, D.K., Ellisman, M.H., and Sejnowski, T.J. (2005). Evidence for ectopic neurotransmission at a neuronal synapse. Science *309*, 446–451.
22. Bray, D., Bourret, R.B., and Simon, M.I. (1993). Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. Mol. Biol. Cell *4*, 469–482.
23. Levin, M.D., Morton-Firth, C.J., Abouhamad, W.N., Bourret, R.B., and Bray, D. (1998). Origin of individual swimming behavior in bacteria. Biophys. J. *74*, 175–181.
24. Barkai, N., Alon, U., and Leibler, S. (2001). Robust amplification in adaptive signal transduction networks. C. R. Acad. Sci. Paris *2*, 1–7.
25. Morton-Firth, C.J., and Bray, D. (1998). Predicting temporal fluctuations in an intracellular signalling pathway. J. Theor. Biol. *192*, 117–128.
26. Rao, C.V., Kirby, J.R., and Arkin, A.P. (2005). Phosphatase localization in bacterial chemotaxis: divergent mechanisms, convergent principles. Phys. Biol. *2*, 148–158.
27. Bray, D., and Duke, T. (2004). Conformational spread: the propagation of allosteric states in large multiprotein complexes. Ann. Rev. Biophys. Biomol. Struct. *33*, 53–73.

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, M.S. 977-152, Berkeley, California 94720, USA. [2]Department of Bioengineering, and Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA.
E-mail: ssandrews@lbl.gov

Correspondences

Elephants avoid costly mountaineering

Jake Wall[1,2], Iain Douglas-Hamilton[1] and Fritz Vollrath[3,4]

Understanding the behavioural decisions underlying animal movements is a major challenge. Here we report evidence for the importance of the abiotic terrain feature 'gradient' in guiding the movements of African savannah elephants (*Loxodonta africana*). Global Positioning System (GPS) tracking data overlaid onto digital elevation and surface gradient models show that elephants tend to avoid steep slopes. Energy calculations suggest that even minor hills are considerable energy barriers for heavy animals.

Elephants are keystone animals in Africa and Asia [1], and effective conservation planning strategies must integrate a thorough knowledge of the range use and spatial requirements of these magnificent animals. Only with such knowledge can we ensure that elephants will be able to survive despite increasingly aggressive human encroachment into their traditional territory [2]. Moreover, there is much to be learned scientifically from understanding the ecological requirements — as well as limitations — of the last remaining representatives of a once cosmopolitan and ecologically critical megafauna.

Early studies of elephant movements deployed radio tracking from the air and provided rather infrequent 'fixes' which painted an incomplete picture of spatial utilisation [3]. Modern GPS collars using a satellite and/or cell-phone link allow us to collect movement data with high temporal and spatial resolution [4], reflecting true range use by also mapping areas not visited. Long-term elephant tracking studies are beginning to show