III. NO FINE TUNING

Imagine making a model of all the chemical reactions that occur inside a cell. Surely this model will have many thousands of variables, described thousands of differential equations. If we write down this many differential equations with the right general form but choose the parameters at random, presumably the resulting dynamics will be chaotic. Although there are periodic spurts of interest in the possibility of chaos in biological systems, it seems clear that this sort of "generic" behavior of large dynamical systems is not what characterizes life. On the other hand, it is not acceptable to claim that everything works because every parameter has been set to just the right value—in particular these parameters depend on details that might not be under the cell's control, such as the temperature or concentration of nutrients in the environment. More specifically, the dynamics of a cell depend on how many copies of each protein the cell makes, and one either has to believe that everything works no matter how many copies are made (within reason), or that the cell has ways of exerting precise control over this number; either answer would be interesting. This problem—the balance between robustness and fine tuning—arises at many different levels of biological organization. Our goal in this chapter is to look at several examples, from single molecules to brains, hoping to see the common themes. [This seems to be the thinnest, and least well worked out of all the four main chapters. All advice is welcome!]

Physics, especially theoretical physics, is the search for concise mathematical descriptions of Nature, and to a remarkable extent this search has been successful. The dirty laundry of this enterprise is that our mathematical descriptions of the world have parameters. In a sense, one mathematical structure describes several possible worlds, which would be somewhat different if the parameters were chosen differently. Sometimes this variety is a good thing—in condensed matter physics, for example, the different parameter values might correspond to genuinely different materials, all of which are experimentally realizable. On the other hand, if the predictions of the model are too sensitive to the exact values of the parameters, there is something vaguely unsatisfying about our claim to have explained things. Such strongly parameter–dependent explanations are often called "finely tuned," and we have grown to be suspicious of fine tuning. Experience suggests that if parameters need to be set to precise (or somehow unnatural) values, then we are missing something in our mathematical description of Nature.[60]

One needs, of course, to be cautious in identifying examples of fine tuning. As an example, many of the beautiful phenomena associated with solar eclipses depend on the fact that, seen from our vantage point on the earth, the angular size of the moon is almost exactly equal to the angular size of the sun. As far as we know, this is a coincidence, and isn't connected to anything else. Presumably this coincidence (which, at certain times of year, occurs with $\sim 1\%$ precision) is related to the fact that there are many planets with moons—even more if we count the planets orbiting other stars—and we happen to live on one of them. Thus, we are sampling one out of many possibilities, and so rare things will happen. Similarly, elections sometimes turn on a surprisingly small number of votes, a tiny fraction of the total. This might seem like some sort of fine tuning,[61] but it is also true that *most* elections do not have outcomes anywhere near the point of perfect balance among the outcomes. This is more obviously one of those cases in which we are sampling many examples, and finely tuned outcomes will happen, sometimes, by chance alone. What we need to worry about are cases in which fine tuning seems essential to make things work (unlike the moon/sun example), and where we see this in representative examples, or in all examples (unlike the elections). We'll see plenty of these problematic cases.

In biological systems, there may be different reasons to be suspicious of fine tuning. On the one hand, for many processes what we call parameters are certainly dynamical variables on longer time scales (such as the number of copies of a protein), and there is widespread doubt that cells can regulate these dynamics precisely. More fundamentally, the parameters of biological systems are encoded in the genome, and in order for evolution to occur it seems necessary that, near to the genomes we see today, there must be genomes (and hence parameter values) which also generate functional organisms of reasonable fitness. These ideas have entered the literature as the need for robustness and evolvability. Note that while the physicist's suspicion of fine tuning is a statement about the kind of explanation that we find satisfying, any attempt to enshrine robustness and evolvability as specifically biological principles involves hypotheses, either about the ability of cells to control their internal states or about the dynamics of evolution.

In this section we will look at several examples of the fine problem, starting at the level of single molecules and then moving "up" to the dynamics of single neurons, the internal states of single cells more generally, and networks

---

[60] At this point I usually try to remind the students of examples—the apparent vanishing of CP violation for the strong interaction, and the prediction of the axion as a solution to this problem, is a favorite. The cosmological constant is another one. Whether these remarks help depends on what the students have learned in other courses. Would it be good to make this explicit here? In the text or a footnote?

[61] We'll leave aside, for this discussion, the disturbing possibility that vote totals *are* being tuned by some process that is separate from the actions of the voters themselves.

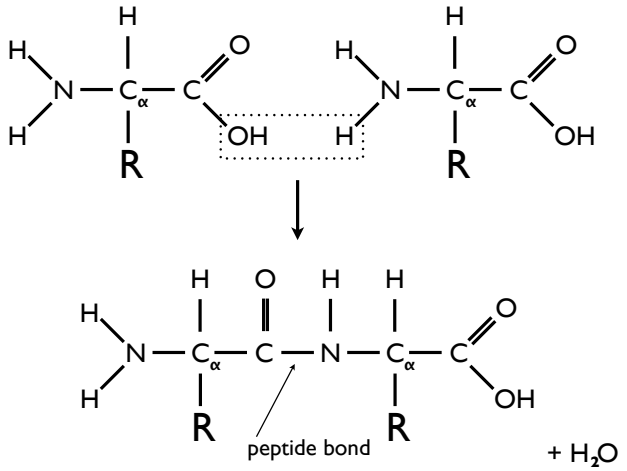

FIG. 79 The basic structure of amino acids and the peptide bond. At top, two amino acids. Different amino acids are distinguished by different groups **R** attached to the α–carbon. Proteins are polymers of amino acids, and the chemical step in polymerization is the formation of the "peptide bond" by removal of a water molecule.

of neurons. As noted at the outset, these different biological systems are the subjects of non–overlapping literatures, and so part of what I hope to accomplish in this Chapter is to highlight the commonality of the physics questions that have been raised in these very different biological contexts.

A. Sequence ensembles

The qualitative ideas about robustness vs fine tuning can be made much more concrete by focusing on single protein molecules. We recall that proteins are heteropolymers of amino acids (Fig 79), each monomer along the polymer chain chosen from twenty possible amino acids (Fig 80). When we look at the proteins made by one particular organism, of course each protein has some particular sequence. If a typical protein is 200 amino acids long, then there are $(20)^{200} \sim 10^{260}$ possible sequences, out of which a bacterium might choose a few thousand, and we choose a few tens of thousands. While different organisms do make slightly different choices, even if we sum over all life forms on earth we will find that real proteins occupy a very small fraction of the available volume in sequence space.

Proteins with different sequences fold up into different structures and carry out different functions. Thus, the sequence obviously matters. Yet, it can't be that the *exact* sequence matters, and this can be checked experimentally. Although some changes are disastrous (e.g., trying to bury a charged amino acid deep in the interior of the protein), many amino acid substitutions leave the structure and function of a protein almost completely unchanged, and many more generate quantitative modulations of function which could be useful in different environments or for closely related organisms. [Should add some figures with protein structures. Need pointer to Appendix A.5 discussing methods of structure determination. Also need to point out that the possible folds seem to be limited, which is another indication that not all details matter.]

Although protein function is tolerant to a wide range of sequence changes, not all sequences really make functional proteins. We can make this statement both as a theoretical result and as an experimental fact. Experimentally, we can synthesize proteins by choosing amino acids at random, and almost none of these will fold. As we will see below, we can even bias our choices at each site, trying to emulate a known family of proteins, and it still is true that if we choose each amino acid independently, most proteins don't fold.

As a crude theoretical model of a protein, we can coarse grain to keep track of the positions $\mathbf{r}_i$ of each α–carbon atom (see Fig 79) along the chain, not worrying about the detailed configuration of the side chains that project from the backbone. Successive amino acids are bonded to one another, with a relatively fixed bond length ℓ, and when the chain folds to bring two amino acids near one another they have an interaction that depends on their identity, plus an excluded volume interaction that is independent of identity. So the total energy looks something like

$$E(\{\mathbf{r}_i\}) = \frac{\kappa}{2}\sum_i \left(|\mathbf{r}_{i+1} - \mathbf{r}_i| - \ell\right)^2 + \frac{1}{2}\sum_{ij} V(S_i, S_j)u(\mathbf{r}_{i+1} - \mathbf{r}_i) + \frac{1}{2}\sum_{ij}\Delta(\mathbf{r}_{i+1} - \mathbf{r}_i), \tag{}$$

where the stiffness κ should be large, the function $u(\mathbf{r})$ needs a shape to express the fact that amino acids have their optimal interaction at finite separation of their centers, and $\Delta(\mathbf{r})$ should be relatively short ranged to express the excluded volume effect. We could try to be a little more realistic and have an extra variable for each amino acid, to keep track of the configuration of the side chain which project from the position $\mathbf{r}_i$.

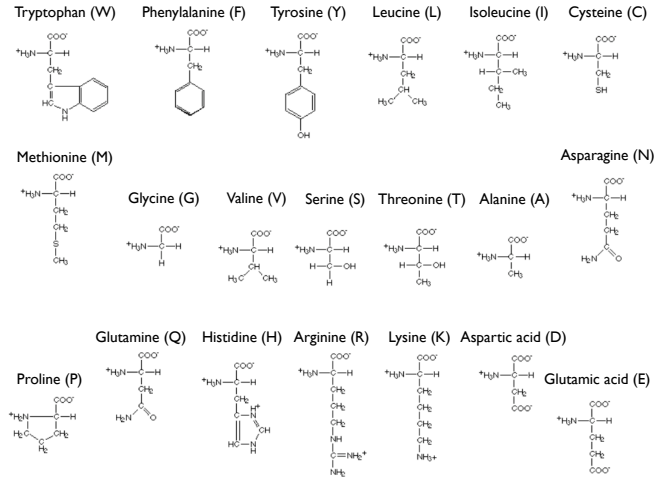

FIG. 80 The twenty different amino acids, arranged from most hydrophobic (top left) to most hydrophilic (bottom right). [perhaps should redraw for better consistency with Fig 79; show only R groups?]

Problem 90: Screening. We are assuming that all interactions extend only over short distances, but we also know that there are charged groups. In this problem you'll show that the long ranged Coulomb interaction is screened. For simplicity, let's imagine that everything is happening in an aqueous solution with only two types of ions, one positive and one negative (e.g., a simple salt solution, where the ions are Na^+ and Cl^-). Let the density of the two ions be $\rho_+(\mathbf{x})$ and $\rho_-(\mathbf{x})$, respectively. If the local electrical potential is $\phi(\mathbf{x})$, then in equilibrium the charge densities must obey

$$\rho_\pm(\mathbf{x}) = \rho_0 \exp\left[\pm\frac{q_e\phi(\mathbf{x})}{k_BT}\right], \tag{475}$$

where q_e is the charge on the electron and ρ_0 is the density or concentration of ions in the absence of fields. Suppose that we introduce an extra charge Z at the origin. Convince yourself that the potential then obeys

$$\nabla^2\phi(\mathbf{x}) = \frac{1}{\epsilon}\left[Zq_e\delta(\mathbf{x}) + q_e[\rho_+(\mathbf{x}) - \rho_-(\mathbf{x})]\right], \tag{476}$$

where ϵ is the dielectric constant. The combination of these two equations is often called the "Poisson–Boltzmann" model, since Eq (475) is the Boltzmann distribution and Eq (476) is the Poisson equation of electrostatics. [I have avoided issues of units in electrostatics until now .. get this straight, because we need numbers at the end!]

(a.) Show that, if the spatial variations in potential are small, Eq's (475) and (476) can be combined to give

$$\nabla^2\phi(\mathbf{x}) + \frac{1}{\lambda^2}\phi(\mathbf{x}) = Zq_e\delta(\mathbf{x}). \tag{477}$$

What is the length λ in terms of the other parameters in the problem?

(b.) You may remember that Eq (477) has solutions that decay exponentially far from the origin; this is the same as for a force mediated by the exchange of a massive particle as opposed to the electromagnetic force, mediated by the massless photon.[62] In this context, Eq (477) is called he Debye–Hückel equation. Solve Eq (477) to give this result explicitly. If the typical concentration of ions in solution is $\rho_0 \sim 100\,\text{mM}$, what is the value of λ?

(c.) With only two univalent ion species, their relatively concentrations are fixed by neutrality, and thus there is only one parameter ρ_0 that enters the discussion. Generalize the derivation of the linearized Eq (477) to the case where there are many species of ions.

(d.) Going back to the two–species case in Eq (476), can you solve the problem without making the linearizing approximation that leads to Eq (477)? With spherical symmetry it's a one dimensional problem, so at worst you should be able to do this numerically. With ρ_0 in the range of 100 mM as above, how good is the linearized theory?

At the end of all this, does it seem reasonable that even electrostatic interactions are effectively local?

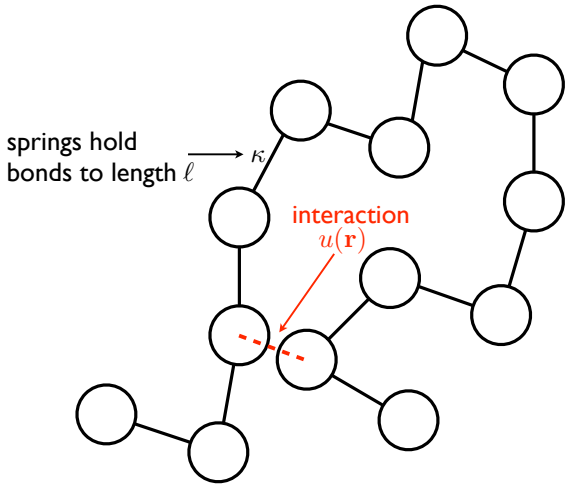

FIG. 81 A model for proteins, after Eq (475). Bonds with stiff springs connect neighboring amino acids, which interact through a potential $u(\mathbf{r})$ when they get close. The strength of the interaction is modulated by the identity of the amino acids through the term $V(S_i, S_j)$ in Eq (475).

If we set the interaction $V = 0$, Eq (475) describes a polymer that takes a self–avoiding random walk. If $V = -V_0$, then there is a net attraction that causes collapse of the polymer into a more compact phase at low temperature, but this state is still disordered, since there is nothing to prefer one compact configuration over another. If V depends on the amino acid identities, then if we choose the sequence at random the effective interaction between monomers i and j will also be random. Although this sounds like a complicated problem, we know a great deal about the behavior of systems where the Hamiltonian contains terms chosen at random.

[62] Historically, this idea goes back to Yukawa, who imagined the strong force between protons and neutrons mediated by the exchange of a heavy particle. We now know that this was on the right track, but there were more layers of the strong interaction to be uncovered; solutions to Eq (477) are still called Yukawa potentials. A more direct connection to the standard model of particle physics is in the case of the weak interaction, where the large mass of the $W^\pm$ and Z bosons are directly related to the short range over which the weak interaction is effective.

The prototype of a system with random interactions is the spin glass. Imagine a solid in which, at every site, there is a magnetic dipole which can point up or down, and hence can be described by an Ising spin $\sigma_\mu = \pm 1$ at site μ. If neighboring spins tend to be parallel, then we can write the Hamiltonian as

$$H = -J \sum_{\langle \mathrm{i,j} \rangle} \sigma_\mathrm{i} \sigma_\mathrm{j}, \tag{478}$$

where $\langle \mathrm{i,j} \rangle$ denotes neighboring sites. In the classic spin glass materials, magnetic impurities are dissolved in a metal, so the distances between neighbors are random. Further, when the conduction electrons in the metal respond to the magnetic impurity, they polarize, but in a metal all the electronic states involved in responses to small perturbations are near the Fermi surface, and hence have a very limited range of momenta or wavevectors in their wavefunctions. This limitation in momentum space corresponds to an oscillation in real space, so the polarization surrounding a single magnetic impurity oscillates with distance; a neighboring impurity will 'feel' this polarization, and so the effective interaction between the two impurities can be positive or negative, at random, depending on the distance between them. This suggests a Hamiltonian of the form

$$H = -\sum_{\mathrm{ij}} J_\mathrm{ij} \sigma_\mathrm{i} \sigma_\mathrm{j}, \tag{479}$$

where J_ij is a random number. In a real system these interactions would be nonzero only for nearby spins, but there is a natural "mean field" limit in which we allow all the spins to interact; this is the Sherrington–Kirkpatrick model.

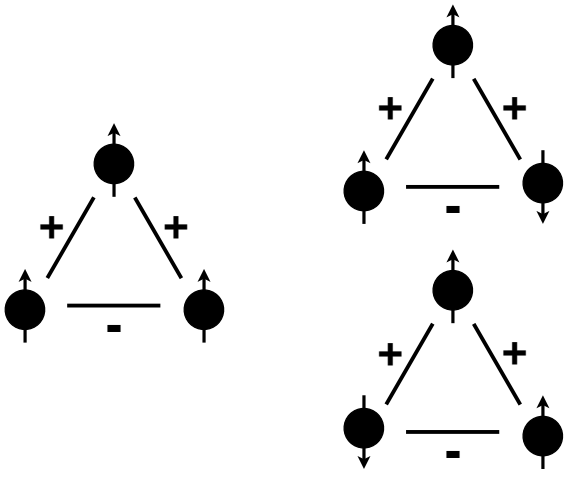

FIG. 82 Three frustrated spins. Signs on the bonds indicate the signs of J_ij in Eq (479). No matter what configuration of spins we choose, one of the bonds is always unsatisfied.

The key qualitative idea in spin glass theory is *frustration*, schematized in Fig 82. In the case of the "ferromagnetic" Ising model in Eq (478), each term in the Hamiltonian can be made as negative as possible by having all the spins point in the same direction, either up or down. But, in the spin glass case, we may find (for example) that spin 1 is coupled to spins 2 and 3 with ferromagnetic interactions $J_{12} > 0$ and $J_{13} > 0$, but spins 2 and 3 are coupled to each other with an anti–ferromagnetic interaction, $J_{23} < 0$. In such a triangle, there is no configuration of the spins which can optimize all the terms in the energy function simultaneously—the interactions compete. As one can see in this simple problem with three spins, a consequence of this competition is that there are many states of the system with low energy that are nearly degenerate. Importantly, in systems with many spins these low lying states correspond to very different spin configurations.

Problem 91: Simulating (small) spin glasses. Consider a mean field spin glass, as in Eq (479), in which the couplings $J_{\mu\nu}$ are drawn at random from a Gaussian distribution; for simplicity start with the assumption that the mean of this distribution is zero and the variance is one. Notice that with N spins there are exactly 2^N states of the system as a whole, so that up to $N = 20$ (or even a bit more) you can easily enumerate all of these states without taxing the memory of your laptop.

(a.) Write a simple program (e.g., in MATLAB) which, starting from a particular random matrix $J_{\mu\nu}$, gives the energies of all the states in an N spin system.

(b.) Find the ground state energy of an N spin system, and do this many times for independent choices of the random interactions $J_{\mu\nu}$. Show that, if the distribution out of which the $J_{\mu\nu}$ are drawn is held fixed, then the ground state energy does not seem to be extensive (i.e., proportional to N) as N varies. In contrast, if the variance of J scales $\propto 1/N$, show that the average ground state energy does seem to be proportional to the number of spins. Can you give an analytic argument for why this scaling should work?

(c.) The exact ground state energy depends on the particular choice of the interactions $J_{\mu\nu}$. One might hope that, as the system becomes large, there is a "self–averaging," so that the energy per spin becomes independent of these details in the limit $N \to \infty$. Do you see any signs of this?

(d.) Having normalized the variance of the couplings $\langle J^2 \rangle = 1/N$, so that the ground state energy is on the order of -1 per spin, compute the gap Δ between the ground state and the first excited state of the system, again for many realizations of the matrix $J_{\mu\nu}$. How does the probability distribution of this gap behave at small values of the gap? In particular, is there a finite probability density as $\Delta \to 0$? How does this behavior of the gap compare with what you expect in a ferromagnet?

(e.) Show that at least some of the low lying states have spin configurations that are very different from the ground state. Again, contrast this with the case of a ferromagnet.

The statistical mechanics of spin glasses is a very beautiful subject, and we could spend a whole semester on this. What we need for the moment, however, is an intuition, something of the sort one can get from the numerical simulation above. In systems with substantial frustration, we expect that there will be many locally stable

low energy states, and these will be very far apart in the relevant state space. Thus, rather than having a well defined ground state, with small fluctuations around this state, there are many inequivalent near–ground states, often with large barriers between them. If we think of the dynamics of the system as motion on an energy surface, then this surface will be rough, with many valleys separated by high passes; indeed, in the Sherrington–Kirkpatrick model there are valleys within valleys, hierarchically. This needs a figure. It's a bit conventional, but maybe there is a reason for the convention?

What does all of this teach us about the protein folding problem? To the extent that we can make analogies between spin glasses and heteropolymers with random sequences, we expect that these randomly chosen proteins will not, in general, have unique ground state structures. Instead, there will be many inequivalent structures with nearly the same low energy, separated by large barriers. Several groups have used modern tools from the statistical mechanics of disordered systems to make this intuition precise Should I say something about the heftier calculations? An Appendix about replicas? Where else do we really need those ideas?], and indeed the random heteropolymer is a kind of glass—the polymer has compact, locally stable structures, but there are many of these, and the system tends to get 'stuck' in one or another such local minimum at random. This contrasts sharply the ability of real proteins to fold into particular, compact conformations that are (at some level of coarse graining) unique, determined by the sequence. The real problem is even worse, because we have only considered the statistical mechanics of one polymer in solution; in practice the folded state of proteins competes not only with the higher entropy unfolded state, but with states in which multiple protein molecules aggregate and precipitate out of solution.

The conclusion is that the proteins which occur in Nature cannot be typical of sequences chosen at random. At the same time, not every detail of the amino acid sequence can be important. This is perhaps the most fundamental example of the general question we are exploring in this Chapter—our description of life cannot depend on fine tuning, but neither are the phenomena of life generic. Concretely, we can ask how to describe the ensemble of sequences that we see in real proteins. One possibility is that this ensemble is profoundly shaped by history, and surely at some level this is true—we can trace evolutionary relationships through sequence data. Another possibility is that the ensemble of *possible* sequences is enormously constrained by physical principles—ensuring that a protein will fold into some compact, reproducible structure is very difficult, and perhaps even enough to explain the dramatically restricted range of sequences and even structures that we observe in real proteins.

At this point we should pause to note that the problem we are formulating is related to, but different from, a much more widely discussed problem. The general question of how protein structure emerges from the underlying amino acid sequence is referred to as the "the protein folding problem." As a practical matter, one might like to predict the three dimensional structure of the folded state, starting only with the sequence. Many approaches to this problem are based not on a physical model for the interactions, but on attempts to generalize from many known examples of sequence/structure pairs. Faced with a particular sequence from Nature, this is can be an extraordinarily effective approach. But it doesn't tell us why some heteropolymers fold into compact, reproducible states, while others do not, and why (presumably) some sequences will never be seen in real organisms. It is this more general version of the question that concerns us here.

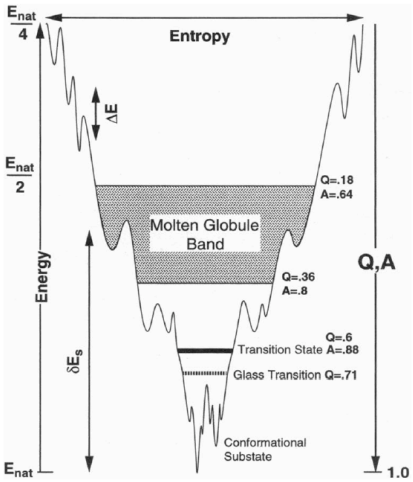

FIG. 83 A schematic energy landscape for protein folding, from Onuchic et al (1995). [Maybe redraw this? Would be good to have equations in the text to point at for features of the funnel.]

One approach emphasizes that in a typical sequence chosen at random, interactions among the different amino acids will be frustrated, blocking the system from finding a single well isolated folded structure of minimum energy. A candidate principle for selecting functional sequences is thus the minimization of this frustration. If frustration is absent, there may be few if any major energetic barriers on the path from an unfolded state to the compact, native conformation, although the need for local structural rearrangements along the path may mean that there is an irreducible 'roughness' to the energy surface that, in a coarse grained picture, will limit the mobility of the system along its path. This scenario has come to be called a folding 'funnel,' emphasizing that there is a single dominant valley in the energy landscape, into which all initial configurations of the system will be drawn, as shown schematically in Fig 83.

At a technical level, if frustration is absent, then we can look at the ground state or native structure and "read off" an approximation to the interactions. Thus, in a ferromagnet, all the spins are parallel in the ground state, and if simply look at each neighboring pair, we would guess that there is a ferromagnetic interaction between them; absent any other data, we should assume that all these interactions have the same strength. Although this might not be exactly right, the Hamiltonian we get in this way will have the correct ground state. In contrast, this doesn't work with spin glasses, because the (near–)ground states necessarily leave some fraction of the interactions unsatisfied, due to frustration. In this spirit, if we look at a small protein, we might try to generate a potential energy function which ties neighboring amino acids together along the chain and, in addition, has "bonds" between amino acids which are in contact in the folded state. We should choose the scale of the potential to have more or less the correct distance between amino acids, and the right order of magnitude for the free energy difference between folded and unfolded states.

Models which bond together amino acids that should form contacts, and neglect all other interactions, actually have a long history, and referred to as Gō models. Concretely, this approach involves an energy function of the form

$$E = \frac{1}{2}\sum_{\text{bonds}} \kappa_r (r - r_0)^2 + \frac{1}{2}\sum_{\text{angles}} \kappa_\theta (\theta - \theta_0)^2 + \frac{1}{2}\sum_{\text{dihedrals}} \sum_n \kappa_\phi^{(n)} \left[1 + \cos(n(\phi - \phi_0))\right]$$
$$+\epsilon \sum_{i<j-3} \left[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^1 2 - C_{ij}^{\text{native}} 6 \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} \right], \tag{480}$$

where the various κs are stiffnesses which hold bond lengths r and angles θ, ϕ along the chain to their native values. The crucial terms are those in the second line, which serve to bond together pairs of residues ij which form a contact in the native, folded state ($C_{ij}^{\text{native}} = 1$) while pushing apart those which do not ($C_{ij}^{\text{native}} = 0$). In principle the different bonds can have specific lengths σ_{ij}, but this is not so important qualitatively.

More recently it has been possible to test these ideas in more detail, by complete simulations of the folding process (cf Fig 84). To summarize the results of the simulation, we can measure the fraction Q of the contacts which should form in the folded state that have actually been made; by construction, as this order parameter increases, the energy of the system decreases. But making contacts lowers the entropy of the polymer, and exactly how much the entropy is lowered depends on which contacts are made. When the dust settles, we can see that the free energy as a function Q has roughly a double well structure. Importantly, one can also sample the configurations in the transition state between the wells, and ask which contacts have been made by the time the molecules finds its way to the top of the barrier. Because there are no competing interactions, the prediction is that the ensemble of transition state configurations must reflect only the geometry of the target, folded state.

Can we test the predictions of such simulations? We expect, from the general arguments in Section II.A, that the rate of folding will have an approximately Arrhenius temperature dependence, $k \propto \exp(-\Delta F/k_B T)$, where ΔF is the free energy difference between the unfolded state and the "transition state" at the top of the barrier.

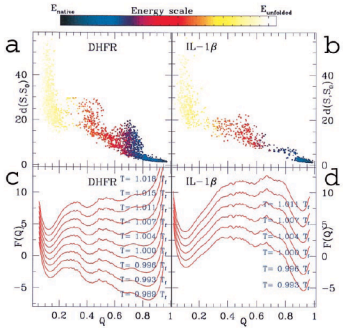

FIG. 84 Gō models for two particular proteins, dihydrofolate reductase (DHFR at left) and interleukin 1β (IL–1β at right), from Clementi et al (2000). Along the x–axis in all figures is a parameter Q measuring the fraction of native contacts that have formed. The top panels show the root–mean–square difference between the structures and the ground state, with colors denoting the energy. Note that, because there are no competing interactions, the energy decreases linearly as more of the native contacts are formed. But different values of Q can be achieved by different numbers of configurations, until at $Q = 1$ there is only one possible structure. Thus the entropy generally declines with Q, although there is also some structure along the way determined by the geometry of the native fold. The result, shown in the bottom panels, is that the free energy has two distinct minima, corresponding to folded ($Q \approx 1$) and unfolded ($Q \approx 0$) states. Different curves correspond to different temperatures, as indicated.

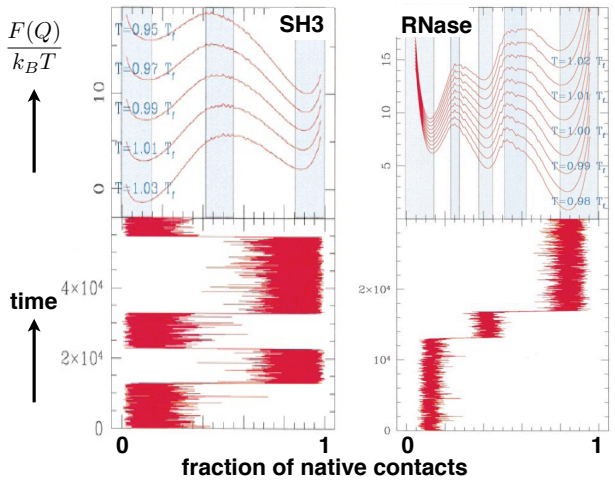

FIG. 85 Simulations of folding for two proteins, using Gō models, from Clementi et al (2000). At each instant of time in the simulation we can count the fraction Q of native contacts, as in Fig 84; sampling the probability distribution of Q we infer the free energy $F(Q)$. At left, simulations of an SH3 domain, which is known to fold rapidly with no obvious intermediate states between folded and unfolded. At right, simulations of the enzyme RNase, which folds more slowly and occupies a well defined intermediate state. These differences are captured by the Gō models, suggesting that frustration does not play a role in slowing the folding of the larger molecules.

Imagine that we mutate the protein to change amino acid i. This has some effect on the free energy of every contact between i and j, and we can measure at least the sum of these effects by measuring the change in the free energy difference between the folded and unfolded states. But if along the "reaction coordinate" Q in Fig 84 these contacts are made (on average) only once $Q > Q_c$, where the Q_c is the position of the transition state, then changing their energy doesn't change the activation free energy for the folding reaction. On the other hand if these contacts are made at $Q < Q_c$, they contribute to the free energy of the transition state and should change the rate of folding. Roughly speaking, the ratio between changes in the (kinetic) free energy of activation and the (thermodynamic) free energy of folding tells us the fraction of contacts involving residue i which are formed in the transition state, and this is something we can get directly from the computations summarized in Fig 84; it is also something one can measure experimentally. Theory and experiment are in surprisingly good agreement [show a figure with the comparison!], which strongly suggests that, at least for small proteins, frustration really has been minimized.

Problem 92: The location of transition states. Suppose that the dynamics of a chemical reaction are described, as in [pointer], by motion of a coordinate x in a potential $V(x)$ that has two minima separated by a barrier. Let the locations of the two minima be at x_1 and x_2, while the peak of the barrier is at a position x_t. Assume that rate constants from transitions between the two wells are governed by the Arrhenius law. Now imagine that we apply a small force f directly to the coordinate x. How does this change the equilibrium between the two states? How does it change the rate of transition, say from the states near x_1 to the states near x_2? Notice that these are measurable quantities. Can you combine them to infer the location of x_t along the line from x_1 to x_2? In particular, can you say something without knowing any additional parameters?

Some proteins are known to fold slowly, moving through a well defined intermediate state. Does this represent a failure to relieve all of the frustration, or is it somehow intrinsic to the size and structure of these molecules? One can make Gō models of thee slower proteins, and compare them with the smaller "two state folders." Results of such a comparison are shown in Fig 85. Perhaps surprisingly, intermediates emerge in the folding of the larger protein even in a model where there is no intrinsic frustration from the interactions among different kinds of amino acids. [I'd like to understand if one be more quantitative here ... can we really conclude that frustration is approximately minimized?]

A second approach to our problem looks more explicitly at the mapping between sequences and structures. The observation that changes in amino acid sequence (mutations) don't necessarily change protein structure tells us that many sequences map into the same structure. But what about the other direction of the mapping? If we imagine some compact structure of a hypothetical protein, can we find a sequence that will fold into this structure? This is the inverse folding problem, or the problem of protein design.

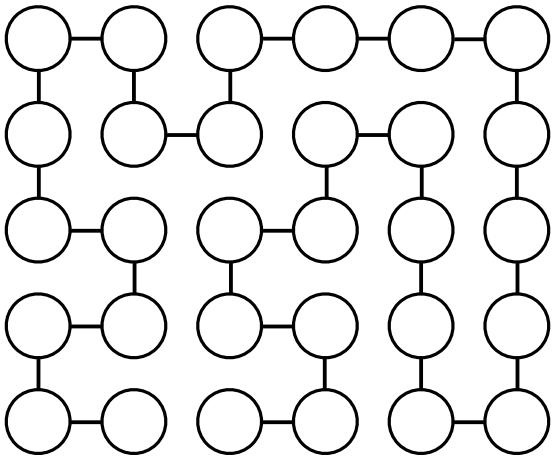

FIG. 86 Compact "folded" structure of an $N = 30$ polymer on a square lattice.

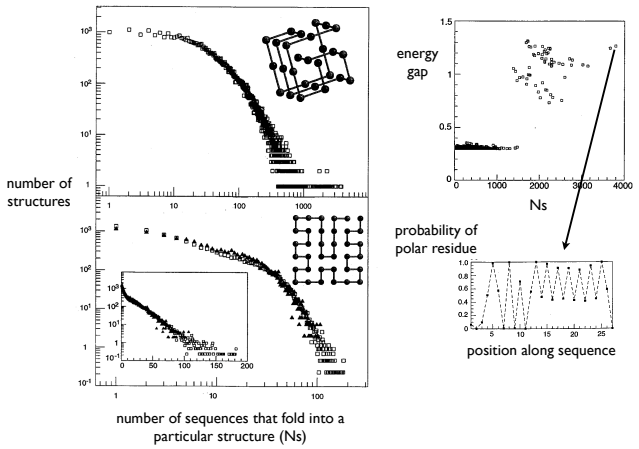

FIG. 87 Exhaustive simulations of compact structures on a lattice, from Li et al (1996). At left, the number of structures which are the ground state for exactly Ns distinct HP sequences, plotted vs Ns for $3 \times 3 \times 3$ (top) and 6×6 (bottom) lattices. Note the small number of structures which are the ground states for huge numbers of sequences. At right, the energy gap between the ground state and the first "excited" state, showing that stability correlates with Ns; the most highly designable structure has a distinctive pattern of hydrophobic and polar residues alternating with residues that are free to be either H or P with nearly equal probability.

To address the inverse folding problem it is helpful to step back and work on a simpler version of the problem. Imagine that there are just two kinds of amino acids, hydrophobic (H) and polar (P). Polar residues are happy to be next to one another, but they are equally happy to on the outside surface of the protein, interacting with water. Hydrophobic residues are much happier to be next to one another, and this includes the effect of not being near water. Finally, for hydrophobic residues, it is likely that having a polar neighbor is marginally better than having water as a neighbor. Thus there are three interaction energies, $E_{PP} > E_{HP} > E_{HH}$, where lower energy is (as usual) more favorable. To simplify yet further, let us assume that the structure of the protein lives on a lattice, as in Fig 86. Now it's clear what we mean by 'compact' structures—if the protein is $N = 27$ amino acids long, for example, a compact structure is one which fills a 3×3 cube—and similarly the definition of 'neighbor' is unambigiuous.

Once we have simplified the problem, it is possible to attack it by exhaustive enumeration. On the $3 \times 3 \times 3$ cube, for example, there are only $\sim 50,000$ inequivalent compact structures, and there are only $2^{27} \sim 10^8$ sequences of this length in the HP model. These numbers are large, but hardly astronomical, so one can explore these sequences and structures completely, also for two dimensional models with $N = 30$ and 36. To begin, out of 2^{27} sequences, less than 5% have a unique compact structure with minimum energy; the majority of sequences have multiple degenerate ground states with inequivalent structures. Conversely, there are nearly 10% of compact structures for which no sequence finds that structure as its ground state; the vast majority of structures are connected to just a handful of sequences. But if we ask how many sequences map into a given structure (N_s), there is a long tail to the distribution of this number (Fig 87, at left), and some structures have thousands of sequences that all reach that structure as their ground state. We can say that these structures are easy to design, or 'highly designable.' Structures with large N_s also have a large energy gap between the compact ground state and the next highest energy conformation, so that highly designable structures are also thermodynamically stable.

What are these highly designable structures? It is hard to extrapolate from such small systems, but certainly the structures with largest N_s have more symmetry and show hints of extended elements such as helices and sheets, as seen in the insets to Fig 87). Can we understand why designability is so variable, and why these particular structures are highly designable?

Before proceeding it is worth noting that finding sequences that stabilize certain structures can be done in two ways. What we really want are sequences with the property that the desired structure is actually the ground state, which means we have to check all other possible competing structures. A weaker notion is to ask for a sequence that assigns a low energy to the desired structure, perhaps even the lowest possible energy across all sequences. If we are just trying the lower the energy, then the problem of choosing sequences is relatively simple—we should try to put the polar residues on the outside, and the hydrophobic residues on the inside. This version of the inverse problem seems at most weakly frustrated, so there are "downhill" paths to find good sequences. [Is there more to say here?]

Analytic approaches to designability describe protein structure not in terms of the positions of all the amino acids, but in terms of a matrix $C_{\rm ij}$ that specifies whether monomers i and j are in contact ($C_{\rm ij} = 1$) or not ($C_{\rm ij} = 0$); by convention $C_{\rm ii} = 0$. Assuming that all long ranged interactions are screened we can approximate the energy of the molecule as having contributions only from amino acids that are in contact,

$$E = \sum_{\rm ij} C_{\rm ij} \sum_{\mu\nu} s_{\rm i}^{\mu} V_{\mu\nu} s_{\rm j}^{\nu}, \qquad (481)$$

where $s_{\rm i}^{\mu} = 1$ if the amino acid at site i is of type μ, and $s_{\rm i}^{\mu} = 0$ otherwise. The matrix $V_{\mu\nu}$ summarizes the interactions among the different types of amino acids. To approach the weaker notion of designability, we need to ask how many sequences give a particular structure a low energy. But asking about the numbers of sequences with a particular energy is just like doing statistical mechanics

where we keep the structure fixed and instead allow the sequence $\{s_i^\mu\}$ to be the dynamical variable. This suggests that we compute the partition function in sequence space,

$$Z_{\rm seq}(C) = \sum_{\{s_i^\mu\}} \exp\left[-\beta \sum_{ij} C_{ij} \sum_{\mu\nu} s_i^\mu V_{\mu\nu} s_j^\nu\right]. \quad (482)$$

Again, this is hard in general, but we can get some intuition by doing a high temperature (small β) expansion.

Summing over all sequences is equivalent to averaging over a distribution in which all sequences are equally likely. Recall that computing the average value of an exponential generates a series of cumulants, or connected correlations:

$$\langle e^{-x}\rangle = \exp\left[-\langle x\rangle + \frac{1}{2}\langle x^2\rangle_c - \frac{1}{3!}\langle x^3\rangle_c + \cdots\right] \quad (483)$$

$$\langle x^2\rangle_c = \langle x^2\rangle - \langle x\rangle^2 = \langle (x-\langle x\rangle)^2\rangle, \quad (484)$$

$$\langle x^3\rangle_c = \langle (x-\langle x\rangle)^3\rangle, \quad (485)$$

and so on. To use this in evaluating $Z_{\rm seq}(C)$, we need to compute quantities of the form

$$\left\langle \sum_{\mu\nu} s_i^\mu V_{\mu\nu} s_j^\nu \right\rangle,$$

or

$$\left\langle \left(\sum_{\mu\nu} s_i^\mu V_{\mu\nu} s_j^\nu \right)^2 \right\rangle.$$

Since we are averaging over a distribution in which all sequences are equally likely, the vector $\vec{s}_i$ that specifies the choice of amino acid at site i is independent of the vectors $\vec{s}_j$ for any $j \neq i$. Pushing through the details, this allows us to show that the free energy

$$F_{\rm seq}(C) \equiv -\frac{1}{\beta}\ln Z_{\rm seq}(C) = A{\rm Tr}(C^2) + B{\rm Tr}(C^3) + \cdots, \quad (486)$$

where the coefficients depend on the details of the potential $V_{\mu\nu}$, and the term $\sim {\rm Tr}(C)$ is absent because ${\rm Tr}(C) = 0$.

Problem 93: Details of $F_{\rm seq}(C)$. Derive Eq (486), carrying the expansion out to at least one more order. Relate the coefficients in the expansion explictly to the properties of the potential $V_{\mu\nu}$.

Because the elements of the matrix C are either 1 or 0, ${\rm Tr}(C^2)$ just counts the number of contacts, while ${\rm Tr}(C^3)$ counts the number of connected paths that lead from site i to site j to site k and back to site i. Similarly, the trace of higher powers counts the number of longer paths. But we can also take a less local view and note that ${\rm Tr}(C^n) = \sum_i \lambda_i^n$, where λ_i are the eigenvalues of the matrix C. As we consider higher powers in the expansion, the result is dominated more and more by the largest of these eigenvalues. Experimenting with small structures as in the discussion above, one can show that the designability of a structure really does correlate strongly with the largest eigenvalue of the contact matrix, and the most designable structure have the largest eigenvalues, as in Fig 88. This is especially interesting since the calculation we have outlined here does not depend on details of the assumptions about the interactions between amino acids—all that matters is locality.

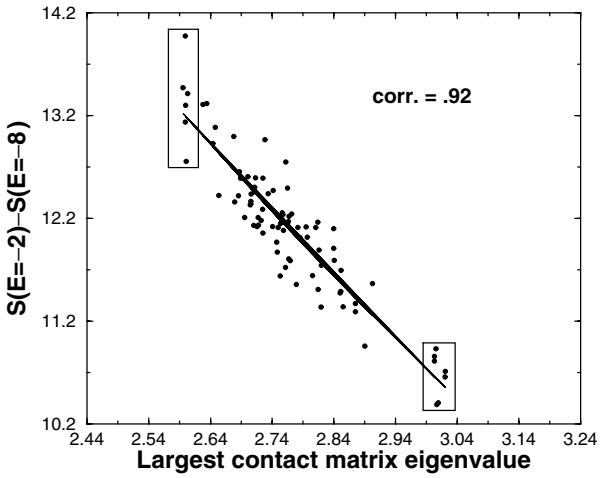

FIG. 88 The connection between designability and the eigenvalues of the contact matrix. [explain]. From England & Shakhnovich (2003).

As noted above, computing $F_{\rm seq}(C)$ gives us a "weak" notion of designability, counting the number of sequences for which a particular structure will have low energy. If we are willing to simplify our model of the interactions, then we can make progress on the stronger notion of designability, that many sequences have the same minimum energy structure. Suppose we return to the model in which there are just two kinds of amino acids, hydrophobic and polar. Further, let's describe the structure in a similar binary fashion, labeling each amino acid by whether it is on the surface of the molecule or in the interior.[63] Now there is a plausible energy function—hydrophobic residues prefer interior sites, polar residues prefer the surface. Thus the energy will be minimized

[63] On a lattice, with the protein folded into a compact structure, this categorization of sites is unambiguous, although one might worry a bit about the more general case.

when the binary description of the sequences ($s_i = +1$ for hydrophobic, $s_i = -1$ for polar) matches the binary description of the structure ($\sigma_i = +1$ for interior, $\sigma_i = -1$ for the surface). Although we might not be able to calculate the exact energy function, ground state structures should correspond to the minimum of a very simple energy that just counts the violations of the hydrophobic/interior, polar/surface rule,

$$E \propto \sum_i (s_i - \sigma_i)^2. \tag{487}$$

An important point about this binary description of structures and sequences is that while all binary strings $\{s_i\}$ represent possible amino acid sequences, not all binary strings $\{\sigma_i\}$ are possible compact structures of a polymer [maybe it would be useful to have a figure illustrating this point?]. Thus in the space of binary strings, and hence H/P sequences, there are special points that correspond to realizable protein structures. The energy function in Eq (487) tells us that the ground state structure for any sequence is the nearest such point, where "near" is measured by a natural metric, the "Hamming distance," counting the number of bits that disagree in the binary string. The set of sequences that will fold into one particular structure are those which fall within the Voronoi polygon surrounding the binary description of that structure, as shown in Fig 89. In this picture, the sequence literally encodes the structure, and the folding process provides a kind of error correction in this code, mapping arbitrary binary strings back to the sparse set of realizable structures. By choosing structures which are far from other structures in this binary representation, one guarantees that many sequences will map to that one structure. Again this picture can be tested against simulations of the lattice models (as in the discussion above), and the results are consistent.

The lesson from all this is that not all structures are created equal, and that selection of structures for their designability induces a nontrivial distribution on the space of sequences. This constraint of course restricts the set of allowed sequences, but at the same time focuses precisely on those sequences for which not all details of the sequence have functional relevance. [check if there is more worth saying here]

There is yet another approach which tries to address the ensemble of allowed sequences, leaning on theory but also using a more direct experimental exploration. In order to appreciate this approach, you need to know that proteins form families. We have already met a simple example of this, with rhodopsin. In your retina, there are four kinds of photoreceptor cells—rods for night vision, and three kinds of cones that provide color vision at higher light intensities—and each one expresses a different pigment molecule, with a different absorption spectrum. Rhodopsin consists of a medium sized organic molecule, retinal embedded in the protein; all the pigments use retinal, so the differences in absorption spectrum reflect differences in the protein. All of these proteins are doing the same job, and have recognizably related structures and amino acid sequences. Nonetheless, they are not identical. In fact, they share sequence and structural similarities with many more proteins, all of which function as receptors (usually for the binding of small molecules rather than the absorption of light), and sit in a membrane rather than being free in solution. Rhodopsin interacts with transducin (Section I.C), which functions as the first stage of an amplification cascade, and other rhodopsin–like molecules interact with similar amplifier molecule. The family to which transducin belongs is called the "G proteins," because part of their function is driven by the hydrolysis of GTP to GDP [be sure this was clear in Chapter 1!], while the rhodopsins and relatives are referred to as G protein coupled receptors (GPCRs). There are GPCRs that respond to hormones, to neurotransmitters in the brain, and, notably, to odorants in the receptor cells of the nose.

Important examples of protein families are provided by enzymes. For example, there are many enzymes which attach phosphate groups to other proteins, for example, and there is variety even within an organism because these protein kinases have different targets; there is even more diversity across organisms. In order to digest our food, we need to cut up the proteins that we ingest, and all cells also need to cut up old proteins that have been damaged or outlived their usefulness in other ways. Cutting the peptide bond quickly and efficiently requires a carefully engineered catalyst, but cells also need control over which sequences they are cutting. Thus there are several families of protein–cutting proteins, called proteases, and there are remarkable structural similarities among molecules separated by billions of yeras of evolu-

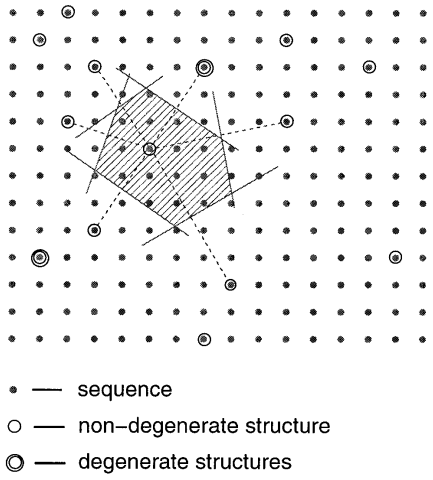

FIG. 89 Designability as seen in the binary description of sequences and structures. [explain]. From Li et al (1998).

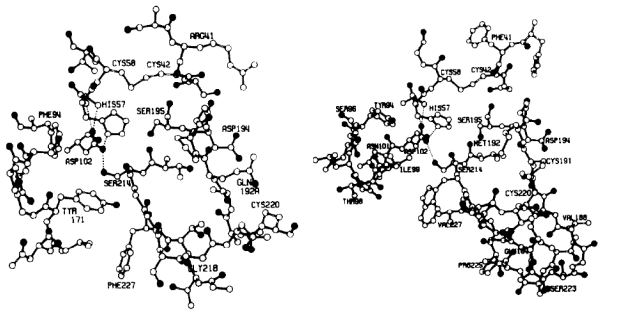

FIG. 90 Comparison of the structure of SGPA (right) and chymotrypsin (left), in the neighborhood of the active site; from Brayer et al (1978). Note in particular the very similar geometrical relations among His57, Asp 102 and Ser 195, the triad of residues involved in the catalytic events.

tionary history. An example is shown in Fig 90, comparing the structure of the bacterial enzyme SGPA and the mammalian enzyme chymotryspin. These molecules have recognizably similar amino acids along only $\sim 25\%$ of their sequences, yet the structures are very similar, especially in the active site where the crucial chemical events occur—the proteins fold to bring these key elements into a very specific geometrical arrangement, despite the sequence differences. Other interesting examples of protein families include smaller parts of proteins (domains) which can fold on their own and function as the interfaces between different molecules; there are hundreds of examples in some of these families.

If we line up the sequences for all the proteins in a family,[64] as in Fig 91, we find that, at each site there are some preferences for one amino acid over another. With enough members in the family, we get a decent estimate of the probability that an amino acid will be chosen in each position along the sequence. Perhaps the simplest hypothesis about the ensemble of allowed sequences is that amino acids are chosen independently at every site, with these probabilities. It should be emphasized that such 'one body' constraints are strong, reducing the entropy of the allowed sequences from a nominal $\sim \log(20)$ per site down to $\sim \log(3)$ per site [check the exact numbers!]. But, this is not enough: if we synthesize proteins at random out of this distribution, it is almost impossible to find one which folds into something like the functional structure characteristic of the original family.

Given that one body models don't work, it seems the next logical step is to look at two body effects: looking across the family of proteins, we see that substitutions at one site tend to be correlated with substitutions at other sites. Can we sample an ensemble of sequences that captures these pairwise correlations? Let us imagine, for simplicity, that there are only two kinds of amino acid; the real case of twenty possibilities just needs more notation. Then we can use $\sigma_i = +1$ for one kind of amino acid at position i, and $\sigma_i = -1$ for the other. The relative frequency of the two choices is measured by the

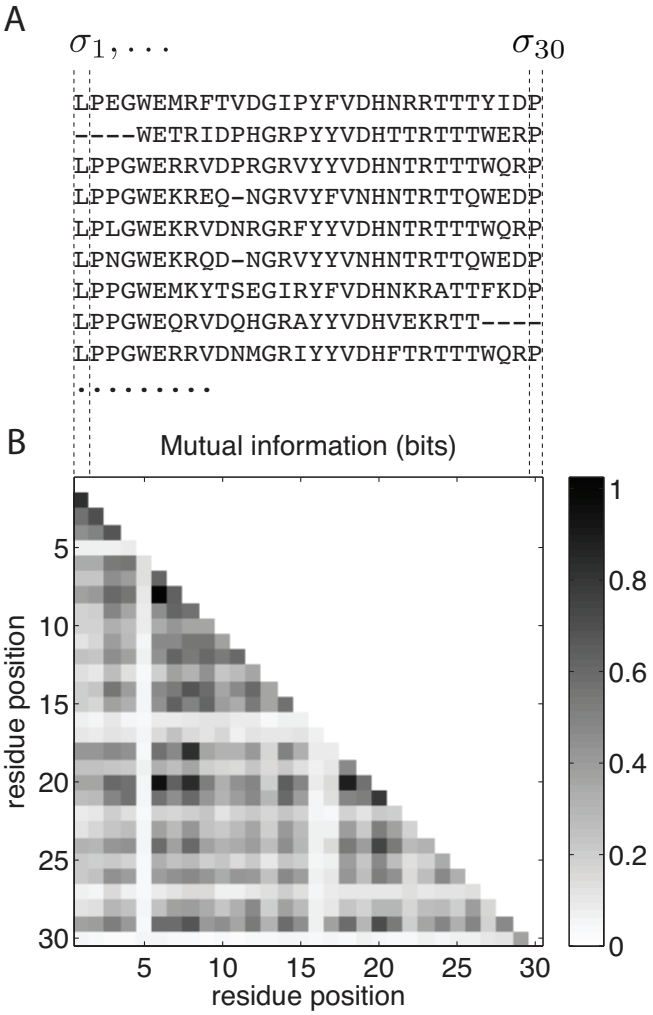

FIG. 91 Alignment of the WW domains, showing (A) the sequences in the family and (B) the correlations between amino acids at pairs of sites, measured by the mutual information. The amino acids are indicated by the one letter codes from Fig 80, with − for gaps. Figure from Mora & Bialek (2011), based on data from [explain source!].

[64] We need to explain that sequence alignment is not trivial. One might even note that algorithms for alignment (or for the recognition of family members) already embody hypotheses about the answer to the question we are trying to formulate here. This all needs some discussion, not least because it points to open problems!

"magnetization" $\langle\sigma_i\rangle_{expt}$, where the subscript remind us that we measure this from data. Similarly, the correlations between amino acid substitutions at pairs of sites is measured by

$$C_{ij}^{expt} \equiv \langle\sigma_i\sigma_j\rangle_{expt} - \langle\sigma_i\rangle_{expt}\langle\sigma_j\rangle_{expt}. \quad (488)$$

Imagine creating an artificial family of M sequences $\{\sigma_i^\mu\}$, with $\mu = 1, 2, \cdots, M$. From this set of replica sequences we can compute the same expectation values that we computed fromt he real family of sequences,

$$\langle\sigma_i\rangle_{model} = \frac{1}{M}\sum_{\mu=1}^{M}\sigma_i^\mu \quad (489)$$

$$C_{ij}^{model} = \frac{1}{M}\sum_{\mu=1}^{M}\sigma_i^\mu\sigma_j^\mu - \langle\sigma_i\rangle_{model}\langle\sigma_j\rangle_{model}. \quad (490)$$

We would like to arrange for the model family of sequences to have these quantities match the experimental ones. The first part ($\langle\sigma_i\rangle_{model} = \langle\sigma_i\rangle_{expt}$) is easy, since we can do this just by choosing the amino acids at every site independently with the same probabilities as in the experimental family. For the two–point correlations, we can form a measure of error between our model sequence ensemble and the real family,

$$\chi^2 = \sum_{ij}\left|C_{ij}^{model} - C_{ij}^{expt}\right|^2, \quad (491)$$

and then we can promote this mean square error to an energy function, and adjust the M sequences according to a Monte Carlo simulation with slowly decreasing (effective) temperature. At low temperatures, this procedure should generate an ensemble of sequences which reproduce the pairwise correlations in the naturally occurring sequences. This procedure has been implemented for a real family of proteins, and novel sequences drawn out of the resulting ensemble have been synthesized. Remarkably, a finite fraction of these sequences fold into something close to the proper native structure, and these folded states are essentially as stable as are the natural proteins. [Reproduce a figure from the Ranganathan work?]

In the limit that we are considering a very large family ($M \to \infty$) of artificial sequences, and we really take the effective temperature to zero, the Monte Carlo procedure draws samples out of a probability distribution that perfectly matches the measured one–point and two–point correlations, but otherwise is as random or unstructured as possible, and hence has maximum entropy. We will meet the maximum entropy idea again in Section III.D, with more details in Appendix A.8. For now, we note that the maximum entropy distribution of sequences takes the form

$$P(\{s_i\}) = \frac{1}{Z}\exp\left[\sum_{i=1}^{N}u_i(s_i) + \frac{1}{2}\sum_{i,j=1}^{N}\mathcal{V}_{ij}(s_i, s_j)\right], \quad (492)$$

where the "fields" u_i and the "interactions" $\mathcal{V}_{ij}$ must be chosen to reproduce the one–point and two–point correlations, where now we allow for the amino acid identity at each site to take on all twenty values, $s_i = 1, 2, \cdots, 20$. Actually finding these fields and interactions is the inverse of the usual problem in statistical mechanics, and can be challenging. But if we can solve this problem, the maximum entropy method provides a potential answer to the question we posed at the outset—if random sequences don't fold, and the exact sequence doesn't matter, how do we describe the ensemble of sequences consistent with a given protein structure or function? Equation (492) gives an explicit answer, a formula for the probability that a particular sequence will occur. Importantly, the form of the distribution is the same as the Boltzmann distribution, with the interactions and fields defining an effective energy surface on the space of sequences. [not sure how to end this .. maybe depends on what Thierry finds in reanalysis of WW domains]

Problem 94: A small maximum entropy model. Give a problem that takes the student through the maxent problem for three spins. Emphasize distinction between interaction and correlation—how much correlation can you get without any direct interactions?

We recall from other problems in statistical mechanics that correlations can extend over much longer distances that the underlying interactions. Thus, although we may detect significant correlations among the amino acid substitutions at many pairs of sites, it is possible that these can be explained by Eq (492) with the interactions $\mathcal{V}_{ij}$ being nonzero only for a very small fraction of pairs ij. Since the physical interactions between amino acids are short ranged, it seems reasonable that if there is a direct connection between the joint choice of residues at sites i and j on the probability that the resulting protein is a member of the family, then sites i and j should be physically close to one another in the protein structure. This idea was worked out in detail for pairs of receptors and associated signaling proteins in bacteria, and it was possible to identify, with high reliability, the amino acids which make up the region of contact between these molecules, as shown in Fig 92. This success raises the tantalizing possibility that we could read off the physical contacts between amino acids—and hence infer the three–dimensional structure of proteins—from analysis of the covariations in amino acid substitutions across a large family.

Should end with some review of what we have learned about the interplay of tuning and robustness; at least some of these questions have become more quantitative.

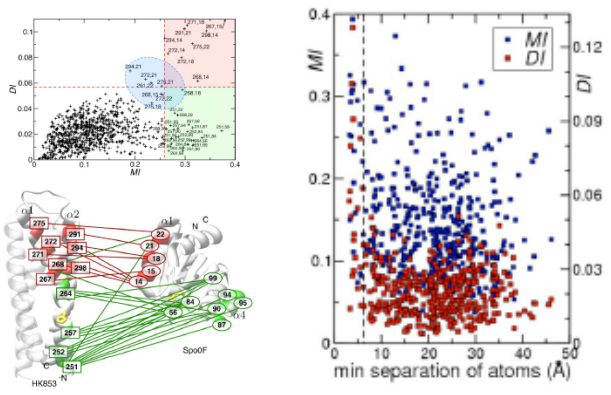

FIG. 92 Interactions between residues in the ensemble of sequences predict spatial proximity, from Weigt et al (2009). [Fill in caption! Do we need more discussion in the text to define "direct information" as generalization of J_{ij}?]

There is also a question about history vs. physics: is the ensemble of sequences *just* a record of evolutionary history, or more like an equilibrium distribution subject to some sensible physical constraints? Do we want to say something explicit about the antibodies? Emphasize that the challenge of building the maximum entropy distributions for larger proteins is really still open?

The amino acid sequences of proteins are translations of the DNA sequences. But there are large parts of DNA which are not coding for proteins. Important parts of this "non–coding" DNA are involved in transcriptional regulation, as discussed in Section II.B. The key steps of this regulatory process involve the binding of transcription factor proteins to DNA, and the architecture of the regulatory network depends on the specificity of these protein–DNA interactions. When we draw an arrow from one transcription factor (TF) to its target gene, then as schematized in Fig [** we had a schematic in a previous chapter, but maybe need another one here?] there must be a short sequence of DNA in or around the target gene to which the transcription factor can bind. The fact that a given TF activates or represses one gene, but not another, then is controlled by the presence or absence of the relevant sequences. But some transcription factors are quite promiscuous, and in higher organisms the relevant sequences often are quite short, so this specificity is not all–or–none. Rather we should think that every short sequence is a possible binding site, and there is a binding energy that depends on the sequence.

Formally, a short piece of DNA sequence can be thought of as a series of bases. Let's write $s_i^\mu = 1$ if the base at position i is of type μ; we have $\mu = 1,\ 2,\ 3,\ 4$ and $i = 1,\ 2,\ \cdots,\ L$, where L is the length of the possible binding site. We can abbreviate $\mathbf{s} \equiv \{s_i^\mu\}$. Then if we look at one transcription factor, there is some binding energy of that factor to the DNA, $E(\mathbf{s})$, for every possible sequence. What does the function $E(\mathbf{s})$ look like? Obviously, if it's a constant then there is no specificity at all—a given transcription factor will influence every gene in the genome—and this can't be right. On the other hand, if the binding is strong only for one specific sequence $\mathbf{s}_0$ (that is, $E(\mathbf{s}) = -E_0$ with large $E_0 > 0$), while $E(\mathbf{s} \neq \mathbf{s}_0) \sim 0$, then the transcription factor can successfully target a small subset of genes, but the landscape for evolutionary change becomes very rugged—changing a single base can completely eliminate one of the regulatory "arrows" in the network, or create a new one of equal strength to all previous arrows—and this doesn't seem right either.

We can turn our question about the form of $E(\mathbf{s})$ around and ask about the set of sequences that will act as functional binding sites, presumably those sequences that have $E(\mathbf{s})$ in some range. In one limit, this ensemble would include all sequences; in the other limit, there would be just one sequence. Thus the issue of specificity in protein–DNA interaction is rather like the problem of amino acid sequence ensembles with which we started this Chapter: where do real biological systems sit along the continuum between completely random sequences at one extreme and unique sequences at the other?

Many of the ideas for analyzing the nature of the sequence ensemble for binding sites involve the starting assumption that each base contributes linearly to the total binding energy, so that

$$E(\mathbf{s}) = \sum_{i=1}^{L} \sum_{\mu=1}^{4} W_{i\mu} s_i^\mu, \tag{493}$$

where $W_{i\mu}$ are the weights given to each position i. One of the first ideas was, in the language we have already used, a maximum entropy argument. If all we know is that functional binding sites must have some average binding energy $\langle E \rangle$, then the maximum entropy distribution consistent with this knowledge is

$$P(\mathbf{s}) = \frac{1}{Z} \exp\left[-\lambda E(\mathbf{s})\right], \tag{494}$$

which of course is the Boltzmann distribution at some effective temperature $\propto 1/\lambda$. Importantly, if the energy is additive as in Eq (493), then the probability of the entire sequence is a product of probabilities at the different sites,

$$P(\mathbf{s}) = \frac{1}{Z} \prod_{i=1}^{L} \exp\left[-\lambda \sum_{\mu=1}^{4} W_{i\mu} s_i^\mu\right]. \tag{495}$$

This means that the expected frequency of occurrence of the different bases at each site—that is, the probability that $s_i^\mu = 1$—can be related directly to the weight matrix,

$$f_{i\mu} \propto \exp\left[-\lambda W_{i\mu}\right]. \tag{496}$$

Thus, if we could get a fair sampling of the ensemble of sequences we could just read off the matrix elements $W_{i\mu}$. [Should I explain that Berg & von Hippel never said "maximum entropy"? Does it matter?]

Problem 95: Random sequences. Take the students through expectations about the distribution of binding energies for the case where sequences are random.

When these ideas first emerged in the mid to late 1980s, in work by Berg & von Hippel, there were few examples where one could point to multiple known binding sites for a single transcription factor. Two important examples were the *lac* operon and the phage λ switch. These are sufficiently important examples in the history of the subject that it is worth taking some time to explain here how they work. [Do this!]

Problem 96: A little more about λ. Depends on what gets said in the text, but maybe ask the students to reproduce Ptashne's argument about the importance of cooperativity.

What was available to Berg and von Hippel were ~ 100 examples of the DNA sequences to which RNA polymerase binds when it begins transcribing. This of course is another example of protein–DNA interaction, not a regulatory interaction but an essential part of all gene expression.[65] Further, there had been in vitro kinetic measurements on transcription, so they knew something about directly about the binding energies. If experiments are done in the regime where the binding sites are usually empty, then the observed transcription rates will be proportional to the concentration of polymerase and the equilibrium constant $K \propto \exp[-\beta E(\mathbf{s})]$. The comparison is shown in Fig 93, including some estimates of errors in the measurements and predictions. The agreement is quite good. Thus, it really does seem that one can, at least roughly, estimate the energetics of binding events from the statistics of sequences, which is quite surprising.

The sequencing of whole genomes, from many organisms, created the opportunity for much more systematic exploration of sequence ensembles. The fact that the number of transcription factors is very much smaller than the number of genes means that, generally, even in a single organism there must be many examples of binding sites for each transcription factor. It seems likely, then, that similar sequences—sequences with good binding energies—will appear more frequently than would be expected at random, and these sequences should, in the simplest cases, be positioned near the start sites of transcription.

In written language, short sequences of letters that occur more frequently than expected by chance have a name—words. When we read, however, there are spaces and punctuation that mark the limits of the words, so we can recognize them. Interestingly, this is less true for spoken language, where the sounds of words often run together, and pauses or gaps are both less distinguishable and less reliable indicators of word boundaries. In fact, we really don't need these markers, even in the case of written text, as you can see by reading Fig 94.

In the simplest view, words are independent, and all structure arises from the fact that not all combinations of

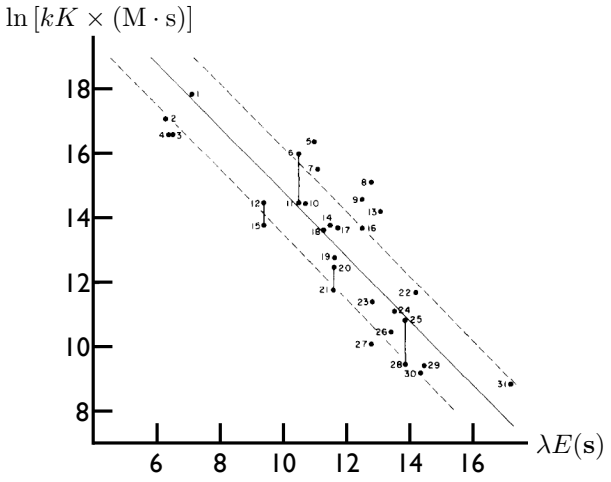

FIG. 93 Sequence dependence of RNA polymerase activity compared with predictions from a maximum entropy model, from Berg & von Hippel (1987). On the vertical axis, effective second–order rate constants for the initiation of transcription by combination of RNA polymerase and different promoter sequences. On the horizontal axis, scaled binding energies predicted from a maximum entropy model based on ~ 100 sequences. Points refer to independent biochemical experiments, with lines connecting measurements on the same sequences, giving a sense for the error bars. A solid line with slope -1 is shown to guide the eye, with dashed lines indicating roughly the errors in the model arising from the finite sample size.

[65] Even in this case the number of sequences is not very large, and we should remember that we are trying to estimate the frequencies of four different bases at each site. To improve their estimates, Berg & von Hippel (1987) used "psuedo–counts," a procedure explained in Appendix A.9.

theresmanalloverforyou
blamingonhisbootsthefa
ultsofhisfeethisisgetting
alarmingoneofthethieves
wassaveditsareasonable
percentagegogo

FIG. 94 A passage from Beckett's *Waiting for Godot*, spoken by Vladimir. All punctuation and spaces have been removed, but (hopefully) the text can still be understood.

letters form legal words. Then, if we know the boundaries between words, the probability of observing a particular text becomes

$$P = \prod_w [P(w)]^{n_w}, \tag{497}$$

where n_w is the number of occurrences of the word w in the text, and $P(w)$ is the probability of this word. But we don't really know, a priori, the correct way of segmenting the text into words, and so we need to sum over all possible segmentations. Each segmentation S generates a different combination of words, so the count $n_w(S)$ depends on S. On the other hand, the probability that a word appears is a property of the language, not of our segmentation, and should be constant. Then

$$P = \sum_S \prod_w [P(w)]^{n_w(S)}. \tag{498}$$

If we think of this as a model for a long text, then given the vocabulary defined by the set of possible words $\{w\}$, maximizing the likelihood of the data amounts to setting the predicted probability of each word to the mean number of occurrences of that word when averaged over all segmentations. Because the text is one–dimensional, there are methods to sum over segmentations that are analogous to transfer function methods for one–dimensional models in statistical mechanics. The real challenge in looking at a genome is that we don't know the vocabulary.

One approach to learning the vocabulary is iterative: start with the assumption that words are single letters, then add two letter words when the frequency of letter pairs is significantly higher than predicted by the model, and so on. To capture the the functional behavior of real biological systems one needs to include words with gaps, such as TTTCCNNNNNNGGAAA, in which "N" can be any nucleotide. Indeed, this example is one of the longer words that emerges from an analysis of possible regulatory regions of the yeast genome, and corresponds to the binding site for MCM1, a protein involved in (among other things) control of the cell cycle. Globally, this approach to "building a dictionary" identifies hundreds words of more than four bases that pass reasonable tests of significance. At the time of the original work, there were $\sim$ 400 known, non–redundant binding sites whose function had been confirmed directly by experiment, and the dictionary reproduced one quarter of these, a success rate 18 standard deviations outside what might have been expected by chance.[66] One can do even better by repeating the analysis using as input text only the regulatory regions of genes whose expression level is affected during particular processes or by the deletion or over–expression of other genes. More power is added to the analysis by using the genomes of closely related organisms. [What do we want to conclude from all of this? Have we lost the notion of binding energy in this discussion?]

Problem 97: Summing over segmentations. Give a problem to connect summing over segmentations with transfer matrix. See Bussemaker et al (2000b).

A very different approach to our problem involves exploring sequence space more systematically. In a relatively short time, several different technologies have emerged for doing this, each of course with its own strengths and weaknesses. [Explain protein binding microarrays, methods from the Quake lab for similar binding measurements, ChiP methods (but chip and seq). Need one good figure illustrating all of these schematically!! Justin provided some input that I haven't digested yet here!]

How do we analyze all these data? Certainly we have the impression that this new generation of experiments provides much more systematic, quantitative data, but there are problems. In the protein binding microarray, for example, there seem to be no reliable calibration of the relation between fluorescence levels and binding probability. Certainly if we see a very bright spot, we can be sure that the protein is bound, but the actual distribution of fluorescence intensities has a long tail, as in Fig 95. Where in this tail do we decide that we have a "hit"?

[66] Say something about what chance means here, and about the general problem of statistical significance in bioinformatics ... .

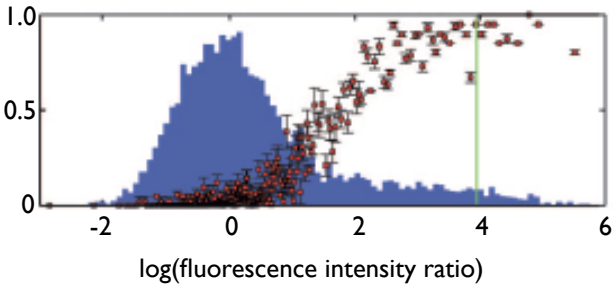

FIG. 95 Protein binding microarray data on the yeast transcription factor Abf1, from Kinney et al (2007). In blue, a histogram of the fluorescence intensities (relative to background) across all $\sim$ 6000 regulatory regions from the yeast genome (Murkherjee et al 2004). In green, the line drawn in the original experiments to define the threshold for binding. In red, with error bars, estimates of the probability that binding has occurred as a function of the fluorescence level, from the analysis described in the text.

In the experiments of Fig 95, fluorescence is a proxy for protein binding, and if things come to equilibrium then this depends on the DNA sequence through the binding energy $E(\mathbf{s})$. The space of sequences is huge, but the model of Eq (493) says that the binding energy is a linear function of the sequence. Thus, fluorescence should depend on sequence only through a single linear projection. Finding this projection is an example of the dimensionality reduction problem discussed in Appendix A.7. The key idea is that, no matter how complicated or noisy the relationship that connects energy to binding to fluorescence, the sequence can't provide more information about the output of the experiment than it does about the more fundamental quantity $E(\mathbf{s})$. Similarly, if we try to summarize the sequence by any reduced description, we will lose information unless our reduction corresponds to estimating $E(\mathbf{s})$ itself. Thus, if we search for a one dimensional description, corresponding to a single linear projections of the sequence that preserves as much information[67] as possible about the experimental output, then the projection we find must be our best linear approximation to $E(\mathbf{s})$, up to a scale factor.[68]

Figure 96 show examples of the weight matrices $W_{i\mu}$ obtained from the "maximally informative dimension" analysis of experiments on the yeast transcription factor Abf1, which is assumed to interact with a 20 base long segment of the DNA. Individual matrix elements typically are determined with better than 10% accuracy, and the interaction of the protein with the DNA evidently is dominated by two approximately symmetric regions of five bases, separated by a gap of another five bases. Importantly, using this method it is possible to analyze in vitro (protein binding microarray) and in vivo (ChiP) experiments, and get consistent answers. In contrast, if we just draw a conservative threshold on the signals strengths (e.g., the green line in Fig 95), then these different sorts of experiments typically lead to divergent interpretations. Once we have confidence in the estimates of $E(\mathbf{s})$, we can go back and ask how the probability that the protein is bound is related to the fluorescence intensity, and this is shown in Fig 95. There is nothing about the analysis that forces this relationship to be smooth or monotonic, but it is.

Can we go further, and relate these linear models of binding energy to the control of gene expression itself? Suppose that we put the expression of a fluorescent protein under the control of a known promoter, and then randomly mutate the sequence. We can then generate an ensemble of bacteria with slightly different sequences, each of which will express the fluorescent protein at different levels, presumably because the relevant transcription factor is binding more or less strongly. Experimen-

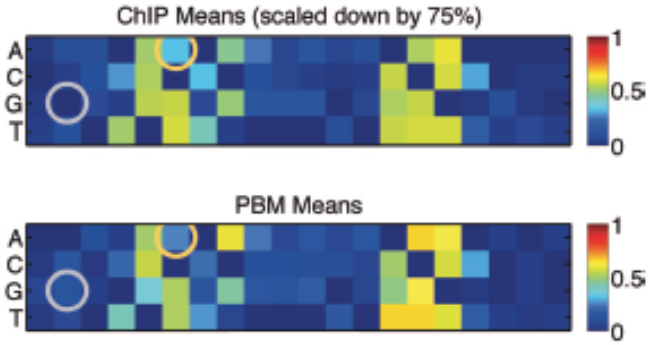

FIG. 96 Weight matrices $W_{i\mu}$ for Abf1 in yeast, from analysis of ChiP (top) and protein binding microarray (bottom) experiments (Kinney et al 2007). In these analyses the overall scale of $E(\mathbf{s})$ is not determined by the data, and so the two results have been scaled to maximize their similarity. Importantly, the two experiments are done in vivo and in vitro, respectively, but nonetheless generate very similar estimates of the underlying matrix governing protein–DNA interactions. The two matrix elements with the poorest agreement are circled, but even these differences have little effect on the predicted binding energies.

[67] "Information" here is used in the technical sense, in bits. See Section IV.A.

[68] The actual computation is a bit more involved because the possible regulatory regions are much larger than the binding sites, and so we have to test not all projections, but all possible projections along the relevant $\sim$ 500 base regions. For details see Kinney et al (2007).

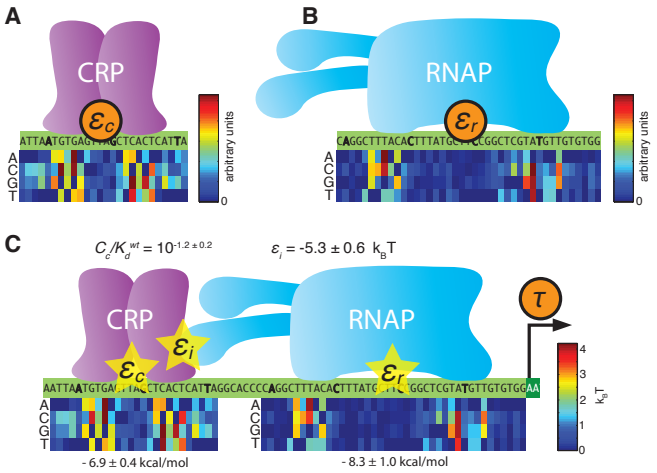

FIG. 97 Analysis of experiments in which the expression of a fluorescent protein is placed under the control of promoter sequences that are randomly mutated versions of the native sequence binding the transcription factor CRP, from Kinney et al (2010). At the top, separate analyses yield the weight matrices $W_{i\mu}$ for the CRP binding site and for the RNA polymerase binding site, up to an arbitrary scale factor. At bottom, a combined analysis places these energies on an absolute scale and determines the interaction energy ϵ_i.

tally, one can sort the cells by their fluorescence, and sequence the promoter regions, and then search once more for a reduction of dimensionality that captures as much information as possible. If the mutations are sprinkled throughout the promoter region, we expect that there are at least two relevant dimensions, corresponding to the binding energy of the transcription factor and the binding energy of the RNA polymerase. The results of such an experiment and analysis are shown in Fig 97.

As before, the search for maximally informative dimensions does not determine the scale of the energies. But if we take seriously that the quantities emerging from the analysis really are energies, then we should be able to compute the probability that the RNA polymerase site is occupied, and it is this occupancy that presumably controls the initiation of transcription. If the energies for binding of the transcription factor (CRP) and RNA polymerase are ϵ_c and ϵ_r, respectively, then the probability of the polymerase site being occupied is

$$\tau = \frac{1}{Z} C_r e^{-\epsilon_r/k_BT} \left(1 + C_c e^{-\epsilon_c/k_BT} e^{-\epsilon_i/k_BT}\right), \quad (499)$$

where the partition function

$$Z = 1 + C_c e^{-\epsilon_c/k_BT} + C_r e^{-\epsilon_r/k_BT} + C_r C_c e^{-\epsilon_r/k_BT} e^{-\epsilon_c/k_BT} e^{-\epsilon_i/k_BT}, \quad (500)$$

where C_c and C_r are the concentrations of the transcription factor and the RNA polymerase, and ϵ_i is the interaction energy between the two proteins when they are both bound to the DNA. Notice that the two binding energies are quantities whose relation to the sequence should already have been determined by search for maximally informative dimensions, except for the scale and zero of energy. By trying to combine these energies we need to set the scale (k_BT) and the zero (equivalently, the concentrations of the proteins), and we have to fit one more parameter, the interaction energy ϵ_i. All of this works, with the results shown at the bottom of Fig 97. For this particular system there are independent measurements of ϵ_i, and there is agreement with $\sim 10\%$ accuracy. Even better, one can show that the single number τ is Eq (499) captures as much information about the sequence dependence of the expression level as do the two numbers ϵ_c and ϵ_r. All of this gives us confidence that the use of statistical mechanics and linear energy models really does make sense here.

Problem 98: RNA polymerase occupancy. Derive Eq (499). Generalize to the case where there are two or more transcription factors, each of which can "touch" the RNA polymerase and contribute an interaction energy. Show that even if the binding of each transcription factor is independent (that is, there are no direct interactions among the TFs), their mutual interactions with the RNA polymerase gives rise to an effective cooperativity in the regulation of transcription. What is the relation of this picture to the MWC models of cooperativity discussed in Appendix A.4?

Now that we have some confidence in our description of the binding energies, we can go back and ask once more about the statistics of sequences, and problem of robustness vs fine tuning. There are several things to say here. I'd like ot cover what happens in Sengupta et al (2002) and Mustonen et al (2008). I think that Justin's observation that you can't find a linear model which points to random collections of genes also is interesting. I'm a bit worried that all of this discussion is in the context of single celled organisms, but there is a lot of stuff to say, e.g., about flies. This needs ALOT of work.

A good general reference about proteins is Fersht (1998). For a modern introduction to polymer physics, see de Gennes (1979). The small simulation in the problems is not a substitute for exploring the theory of spin glasses; the classic papers are collected, with an introduction, by Mézard et al (1986), and a textbook account is given by De Dominicis & Giardina (2006). Early efforts to apply these methods to the random heteropolymer were made by Shakhnovich & Gutin (1989).

De Dominicis & Giardina 2006: *Random Fields and Spin Glasses* C De Dominicis & I Giardina (Cambridge University Press, Cambridge, 2006).

Fersht 1998: *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* AR Ferhst (WH Freeman, San Francisco, 1998).

de Gennes 1979: *Scaling Concepts in Polymer Physics* PG de Gennes (Cornell University Press, Ithaca, 1979).

Mézard et al 1986: *Spin Glass Theory and Beyond* M Mézard, G Parisi & MA Virasoro (World Scientific, Singapore, 1986).

Shakhnovich & Gutin 1989: Formation of unique structure in polypeptide chains: Theoretical investigation with the aid of a replica approach. EI Shakhnovich & AM Gutin, *Biophys Chem* **34,** 187–199 (1989).

Models which incorporate only native interactions, with no frustration, have their origin in work by Gō, reviewed in Gō (1983). A more explicit discussion of minimizing frustration as a principle was given by Bryngelson & Wolynes (1987), and the funnel landscape of Fig 83 is from Onuchic et al (1995). Detailed simulations based on the Gō model are described by Clementi et al (2000a,b).

Bryngleson & Wolynes 1987: Spin glasses and the statistical mechanics of protein folding. JD Bryngelson & PG Wolynes, *Proc Nat'l Acad Sci (USA)* **84,** 7524–7528 (1987).

Clementi et al 2000a: How native-state topology affects the folding of dihydrofolate reductase and interleukin–1β. C Clementi, PA Jennings & JN Onuchic, *Proc Nat'l Acad Sci (USA)* **97,** 5871–5876 (2000).

Clementi et al 2000b: Topological and energetic factors: What determines the structural details of the transition state ensemble and "en–route" intermediates for protein dolding? An investigation for small globular proteins. C Clementi, H Nymeyer & JN Onuchic, *J Mol Biol* **298,** 937–953 (2000).

Gō 1983: Theoretical studies of protein folding. N Gō, *Ann Rev Biophys Bioeng* **12,** 183–210 (1983).

Onuchic et al 1995: Toward an outline of the topography of a realistic protein–folding funnel. JN Onuchic, PG Wolynes, Z Luthey–Schultern & ND Socci, *Proc Nat'l Acad Sci (USA)* **92,** 3626–3630 (1995).

The lattice simulations which explored protein designability were by Li et al (1996). The analytic argument connecting designability to the eigenvalues of the contact matrix was given by England & Shakhnovich (2003), and Li et al (1998) gave the argument relating folding to error correction in the HP model. [Probably there is more to say here!]

England & Shakhnovich 2003: Structural determinant of protein designability. JL England & EI Shakhnovich, *Phys Rev Lett* **90,** 218101 (2003).

Li et al 1996: Emergence of preferred structures in a simple model of protein folding. H Li, R Helling, C Tang & N Wingreen, *Science* **273,** 666–669 (1996).

Li et al 1998: Are protein folds atypical? H Li, C Tang & NS Wingreen, *Proc Nat'l Acad Sci (USA)* **95,** 4987–4990 (1998).

[Need to start with a general reference about protein families] The idea of protein families was essential in the experiments that searched for, and found, the receptors in the olfactory system (Buck & Axel 1991, Axel 2005, Buck 2005). [should give general reference for serine proteases] The structural correspondence between bacterial serine proteases and their mammalian counterparts is from Brayer et al (1978, 1979) and Fujinaga et al (1985). Experiments on the sampling of sequence space while preserving one–point and two–point correlations were done by Socolich et al (2005) and by Russ et al (2005). The equivalence of these ideas to the maximum entropy method was shown in Bialek & Ranganthan (2007). For more on maximum entropy approaches to sequence ensembles, see Weigt et al (2009), Halabi et al (2009), and Mora et al (2010). For a broader view of maximum entropy models applied to biological systems, see Appendix A.8 and Mora & Bialek (2011).

Axel 2005: Scents and sensibility: A molecular logic of olfactory perception. R Axel, in *Les Prix Nobel 2004*, T Frängsmyr, ed, pp 234–256 (Nobel Foundation, Stockholm, 2004).

Bialek & Ranganthan 2007: Rediscovering the power of pairwise interactions. W Bialek & R Ranganathan, arXiv:0712.4397 [q–bio.QM] (2007).

Brayer et al 1979: Molecular structure of crystalline *Streptomyces gresius* protease A at 2.8 Å resolution: II. Molecular conformation, comparison with α–chymotrypsin, and active–site geometry. GD Brayer, LTJ Delbaere & MNG James, *J Mol Biol* **124,** 261–283 (1978).

Brayer et al 1979: Molecular structure of the α–lytic protease from *Myxobacter 495* at 2.8 Å resolution. GD Brayer, LTJ Delbaere & MNG James, *J Mol Biol* **131,** 743–775 (1979).

Buck & Axel 1991: A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. L Buck & R Axel, *Cell* **65,** 175–187 (1991).

Buck 2005: Unraveling the sense of smell. LB Buck, in *Les Prix Nobel 2004*, T Frängsmyr, ed, pp 267–283 (Nobel Foundation, Stockholm, 2004).

Fujinaga et al 1985: Refined structure of α–lytic protease at 1.7 Å resolution: Analysis of hydrogen bonding and solvent structure. M Fujinana, LTJ Delbaere, GD Brayer & MNG James, *L Mol Biol* **183,** 479–502 (1985).

Halabi et al 2009: Protein sectors: Evolutionary units of three–dimensional structure. N Halabi, O Rivoire, S Leibler & R Ranganathan, *Cell* **138,** 774–786 (2009).

Mora & Bialek 2011: Are biological systems poised at criticality? T Mora & W Bialek. *J Stat Phys* **144,** 268–302 (2011); arXiv:1012.2242 [q–bio.QM] (2010).

Mora et al 2010: Maximum entropy models for antibody diversity. T Mora, AM Walczak, W Bialek & CG Callan, *Proc Nat'l Acad Sci (USA)* **107,** 5405–5410 (2010).

Russ et al 2005: Natural–like function in artificial WW domains. WP Russ, DM Lowery, P Mishra, MB Yaffe & R Ranganathan, *Nature* **437,** 579–583 (2005).

Socolich et al 2005: Evolutionary information for specifying a protein fold. M Socolich, SW Lockless, WP Russ, H Lee, KH Gardner & R Ranganathan, *Nature* **437,** 512– 518 (2005).

Weigt et al 2009: Identification of direct residue contacts in protein–protein interaction by message passing. M Weigt, RA White, H Szurmant, JA Hoch & T Hwa, *Proc Nat'l Acad Sci (USA)* **106,** 67–72 (2009).

Should really give some pointers to the problem of sequence alignment!

[Check this against discussion and references in relevant part of Chapter Two!] The modern picture of transcriptional regulation traces its origins to Jacob & Monod (1961), another of the great and classic papers that still are rewarding to read decades after they were published. Their views were motivated primarily by studies of the *lac* operon, and the origins of these reach back to Monod's thesis (1942), which was concerned the phenomenology of bacterial growth. As recounted in Judson (1979), for example, the idea that genes turn on because of the release from repression was due to Szilard; the written record of these ideas is not as clear as it could be, but one can try Szilard (1960). For a modern view, faithful to the history, see Müller–Hill (1996). The other "simple," paradigmatic example of protein–DNA interactions in the regulation of gene expression is the case of bacteriophage λ, which is reviewed by Ptashne (1986), which has also evolved with time (Ptashne 1992); see also Ptashne (2001). These systems provided the background for the pioneering discussion of sequence specificity in protein–DNA interactions (von Hippel & Berg 1986, Berg & von Hippel 1987, 1988). In parallel to this statistical approach, there were direct biochemical measurements of binding energies, and an early attempt to bring these different literatures into correspondence was by Stormo & Fields (1998).

Berg & von Hippel 1987: Selection of DNA binding sites by regulatory proteins. I: Statistical-mechanical theory and application to operators and promoters. OG Berg & PH von Hippel, *J Mol Biol* **193,** 723–743 (1987).

Berg & von Hippel 1988: Selection of DNA binding sites by regulatory proteins. II: The binding specificity of cyclic AMP receptor protein to recognition sites. OG Berg & PH von Hippel, *J Mol Biol* **200,** 709–723 (1988).

von Hippel & Berg 1986: On the specificity of DNA–protein interactions. PH von Hippel& OG Berg, *Proc Nat'l Acad Sci (USA)* **83,** 1608–1612 (1986).

Jacob & Monod 1961: Genetic regulatory mechanisms in the synthesis of proteins. F Jacob & J Monod, *J Mol Biol* **3,** 318–356 (1961).

Judson 1979: *The Eighth Day of Creation* HF Judson (Simon and Schuster, New York, 1979).

Monod 1942: *Recherche sur la Croissance des Cultures Bactériennes* J Monod (Hermann, Paris, 1942).

Müller–Hill 1996: *The lac Operon: A Short History of a Genetic Paradigm* B Müller–Hill (W de Gruyter & Co, Berlin, 1996).

Ptashne 1986: *A Genetic Switch: Gene Control and Phage λ.* M Ptashne (Cell Press, Cambridge MA, 1986).

Ptashne 1992: *A Genetic Switch, Second Edition: Phage λ and Higher Organisms.* M Ptashne (Cell Press, Cambridge MA, 1992).

Ptashne 2001: *Genes and Signals.* M Ptashne (Cold Spring Harbor Laboratory Press, New York, 2001).

Stormo & Fields 1998: Specificity, free energy and information content in protein–DNA interactions. GD Stormo & DS Fields, *Trends Biochem Sci* **23,** 109–113 (1998).

Szilard 1960: The control of the formation of specific proteins in bacteria and in animal cells. L Szilard, *Proc Nat'l Acad Sci (USA)* **46,** 277–292 (1960).

The emergence of whole genome sequences opened several new approaches to the problem of specificity. One important idea is that sequences that are targets for protein binding should have a non–random structure, and we should be able to find this in a relatively unsupervised fashion (Busemaker et al 2000a,b). [Need more here!]

Bussemaker et al 2000a: Building a dictionary for genomes: Identification of presumptive regulatory sites by statistical analysis. H Bussemaker, H Li & ED Siggia, *Proc Nat'l Acad Sci (USA)* **97,** 10096–10100 (2000).

Bussemaker et al 2000b: Regulatory element detection using a probabilistic segmentation algorithm. H Bussemaker, H Li & ED Siggia, *Proc Int Conf Intell Sys Mol Biol* —bf 8, 67–74 (2000).

Need pointers to different large scale experimental approaches—protein binding arrays (Mukherjee et al 2004), ChiP, etc.. Circle back to work from Quake group (Maerkl & Quake 2007). For an approach to the analysis of such measurements making explicit use of dimensionality reduction methods (Appendix **), see Kinney et al (2007). This approach inspired experiments aimed at wider exploration of sequence space (Kinney et al 2010). For other such explorations, see Ligr et al (2006) and Gertz et al (2009).

Gertz et al 2009: Analysis of combinatorial cis–regulation in synthetic and genomic promoters. J Gertz, ED Siggia & BA Cohen, *Nature* **457,** 215–218 (2009).

Kinney et al 2007: Precise physical models of protein–DNA interaction from high-throughput data. JB Kinney, G Tkačik & CG Callan Jr, *Proc Natl Acad Sci (USA)* **104,** 501–506 (2007).

Kinney et al 2010: Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. JB Kinney, A Murugan, CG Callan Jr & EC Cox, *Proc Nat'l Acad Sci (USA)* **107,** 9158–9163 (2010).

Ligr et al 2006: Gene expression from random libraries of yeast promoters. M Ligr, R Siddharthan, FR Cross & ED Siggia, *Genetics* **172,** 2113–2122 (2006).

Maerkl & Quake 2007: A systems appraoch to measuring the binding energy landscape of transcription factors. SJ Maerkl & SR Quake, *Science* **315,** 233–237 (2007).

Mukherjee et al 2004: Rapid analysis of the DNA binding specificities of transcription factors with DNA microarrays. S Mukherjee, MF Berger, G Jona, XS Wang, D Muzzey, M Snyder, RA Young & ML Bulyk, *Nature Genetics* **36,** 1331–1339 (2004).

Need to segue to discussions of evolvability etc.. Probably more references to cite!

Maerkl & Quake 2009: Experimental determination of the evolvability of a transcription factor. SJ Maerkl & SR Quake, *Proc Nat'l Acad Sci (USA)* **106,** 18650–18655 (2006).

Mustonen et al 2008: Energy–depdendent fitness: A quantitative model for the evolution of yeast transcription factor binding sites. V Mustonen, J Kinney, CG Callan Jr & M Lässig, *Proc Nat'l Acad Sci (USA)* **105,** 12376–12381 (2008).

Sengupta et al 2002: Specificity and robustness in transcription control networks. A Sengupta, M Djordjevic & BI Shraiman, *Proc Nat'l Acad Sci (USA)* **, 99,** 2072–2077, (2002).

B. Ion channels and neuronal dynamics

The functional behavior of neurons involves the generation and processing of electrical signals. The dynamics of these currents and voltages are determined by the ion channels which sit in the cell membrane. As noted in our discussion of the rod photoreceptor cell (Section I.C), the cell membrane itself is insulating, and hence there would be no interesting electrical dynamics if not for specific conducting pores. These pores are protein molecules that can change their structure in response to various signals, including the voltage across the membrane, and this means that the system of channels interacting with the voltage constitutes a potentially complex nonlinear dynamical system. We can also think of the ion channels in the cell membrane as a network of interacting protein molecules, with the interactions mediated through the transmembrane voltage. In contrast to many other such biochemical systems, we actually know the equations that describe the network dynamics, and as a result the questions of fine tuning vs. robustness can be posed rather sharply.

When we move from thinking about individual neurons to thinking about circuits and networks of neurons,

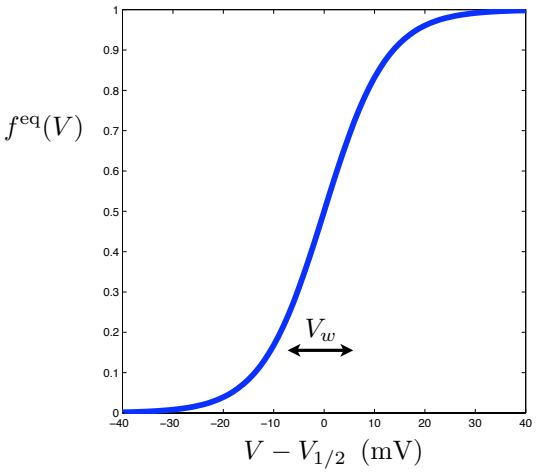

FIG. 98 Activation curve for an ion channel, from Eq (505), with $Q = 4$.

which really do the business of the brain, it is easy to imagine that the neurons are 'circuit elements' with some fixed properties. We enhance this tendency by drawing circuit diagrams in which we keep track of whether neurons excite or inhibit one another, but nothing else about their dynamics is made explicit. In fact, our genome encodes $\sim 10^2$ different kinds of channels, each with its own kinetics, and this range is expanded even further by the fact that many of these channels have multiple subunits, and it is possible to splice together the subunits in different combinations. On the one hand, this creates enormous flexibility, and presumably adds to the computational power of the nervous system. On the other hand, this range of possibilities raises a problem of control. A typical neuron might have eight or nine different kinds of channels, and we will see that the dynamics of the cell depend rather sensitively on how many of each kind of channel is present. In keeping with the theme of this Chapter, it might seem that cells need to tune their channel content very precisely, yet this needs to happen in a robust fashion.

To explore the tradeoff between fine tuning and robustness in neurons, we need to understand the dynamics of the channels themselves. For simplicity, let's neglect the spatial structure of the cell and assume we can talk about a single voltage difference V between inside and outside. Then since the membrane acts as a capacitor, we can write, quite generally,

$$C\frac{dV}{dt} = I_{\rm channels} + I_{\rm ext}, \tag{501}$$

where $I_{\rm ext}$ is any external current that is being injected (perhaps by us as experimenters) and $I_{\rm channels}$ is the current flowing through the channels. Each channel acts more or less as an Ohmic conductance, and the structure of the channel endows it with specificity for particular ions. Since the cell works to keep the concentrations of ions different on the inside and outside of the cell, the thermodynamic driving force for the flow of current includes both the electrical voltage and a difference in chemical potential; it is conventional to summarize this by the "reversal potential" $V_{\rm i}$ for the currents flowing through channels of type i, which might involve a mix of ions. Since current only flows through open channels, we can write

$$I_{\rm channels} = -\sum_{\rm i} g_{\rm i} N_{\rm i} f_{\rm i} (V - V_{\rm i}), \tag{502}$$

where $g_{\rm i}$ is the conductance of one open channel of type i, $N_{\rm i}$ is the total number of these channels, $f_{\rm i}$ is the fraction which are open, and $V_{\rm i}$ is the reversal potential. If each channel has just two states, open and closed, then their dynamics would be described by

$$\frac{df_{\rm i}}{dt} = -\frac{1}{\tau_{\rm i}(V)}\left[f_{\rm i} - f_{\rm i}^{\rm eq}(V)\right]. \tag{503}$$

The equilibrium fraction of open channels as a function of voltage, $f_{\rm i}^{\rm eq}(V)$, often is called the activation curve, and $\tau_{\rm i}(V)$ is the time constant for relaxation to this equilibrium.

What is a reasonable shape for the activation curve? We are describing a protein molecule that can exist in two states, and the equilibrium between these two states depends on voltage. This is possible only if the transition from closed to open rearranges the charges in the protein. In the simplest model, then, the opening of the channel effectively moves a charge Q across the membrane, and so the free energy difference between open and closed states will be $\Delta F = F_0 - QeV$. Then the equilibrium probability of a channel being open will be given by

$$f^{\rm eq}(V) = \frac{1}{1 + \exp\left[(F_0 - QeV)/k_BT\right]} \tag{504}$$

$$= \frac{1}{1 + \exp\left[-(V - V_{1/2})/V_w\right]}, \tag{505}$$

where the point of half maximal activation is $V_{1/2} = F_0/(Qe)$, and the width of the activation curve is $V_w = k_BT/Qe$, as shown in Fig 98. The charge Q is referred to as the "gating charge." We recall that, at room temperature, $k_BT/e = 25\,{\rm mV}$, so that even with relatively small values of Q we expect channels to make the transition from closed to open in a window of $\sim 10\,{\rm mV}$ or so. The location of the midpoint $V_{1/2}$ depends on essentially all aspects of the protein structure in the open and closed states, so it is harder to get intuition for this parameter. In practice, different channels have $V_{1/2}$ values in the range [look this up to give a meaningful survey ..].

It's useful to think about the linearized dynamics; we imagine that there is some steady state at a "resting potential" $V = V_0$, and study small perturbations around

this steady state. The full dynamics are

$$C\frac{dV}{dt} = -\sum_{\mathrm{i}} g_{\mathrm{i}} N_{\mathrm{i}} f_{\mathrm{i}}(V - V_{\mathrm{i}}) + I_{\mathrm{ext}}, \tag{506}$$

$$\frac{df_{\mathrm{i}}}{dt} = -\frac{1}{\tau_{\mathrm{i}}(V)}\left[f_{\mathrm{i}} - f_{\mathrm{i}}^{\mathrm{eq}}(V)\right], \tag{507}$$

and the linearization is

$$C\frac{d\delta V}{dt} = -\sum_{\mathrm{i}} g_{\mathrm{i}} N_{\mathrm{i}} f_{\mathrm{i}}^{\mathrm{eq}}(V)\delta V - \sum_{\mathrm{i}} g_{\mathrm{i}} N_{\mathrm{i}}(V_0 - V_{\mathrm{i}})\delta f_{\mathrm{i}} + I_{\mathrm{ext}}, \tag{508}$$

$$\frac{d\delta f_{\mathrm{i}}}{dt} = -\frac{1}{\tau_{\mathrm{i}}(V_0)}\left[\delta f_{\mathrm{i}} - \left.\frac{df_{\mathrm{i}}^{\mathrm{eq}}(V)}{dV}\right|_{V=V_0}\delta V\right]. \tag{509}$$

Fourier transforming, we can solve for the channel dynamics,

$$\frac{d\delta f_{\mathrm{i}}}{dt} = -\frac{1}{\tau_{\mathrm{i}}(V_0)}\left[\delta f_{\mathrm{i}} - \left.\frac{df_{\mathrm{i}}^{\mathrm{eq}}(V)}{dV}\right|_{V=V_0}\delta V\right] \tag{510}$$

$$-i\omega\delta\tilde{f}_{\mathrm{i}}(\omega) = -\frac{1}{\tau_{\mathrm{i}}(V_0)}\left[\delta\tilde{f}_{\mathrm{i}}(\omega) - \left.\frac{df_{\mathrm{i}}^{\mathrm{eq}}(V)}{dV}\right|_{V=V_0}\delta\tilde{V}(\omega)\right] \tag{511}$$

$$\delta\tilde{f}_{\mathrm{i}}(\omega) = \frac{[df_{\mathrm{i}}^{\mathrm{eq}}(V)/dV]_0}{-i\omega + 1/\tau_{\mathrm{i}}(V_0)}\delta\tilde{V}(\omega), \tag{512}$$

and then substitute,

$$C\frac{d\delta V}{dt} = -\sum_{\mathrm{i}} g_{\mathrm{i}} N_{\mathrm{i}} f_{\mathrm{i}}^{\mathrm{eq}}(V)\delta V - \sum_{\mathrm{i}} g_{\mathrm{i}} N_{\mathrm{i}}(V_0 - V_{\mathrm{i}})\delta f_{\mathrm{i}} + I_{\mathrm{ext}}$$

$$-i\omega C\delta\tilde{V}(\omega) = -\sum_{\mathrm{i}} g_{\mathrm{i}} N_{\mathrm{i}} f_{\mathrm{i}}^{\mathrm{eq}}(V)\delta\tilde{V}(\omega) - \sum_{\mathrm{i}} g_{\mathrm{i}} N_{\mathrm{i}}(V_0 - V_{\mathrm{i}})\delta\tilde{f}_{\mathrm{i}}(\omega) + \tilde{I}_{\mathrm{ext}}(\omega) \tag{513}$$

$$-i\omega C\delta\tilde{V}(\omega) = -\sum_{\mathrm{i}} g_{\mathrm{i}} N_{\mathrm{i}} f_{\mathrm{i}}^{\mathrm{eq}}(V)\delta\tilde{V}(\omega) - \sum_{\mathrm{i}} \frac{[g_{\mathrm{i}} N_{\mathrm{i}}(V_0 - V_{\mathrm{i}}) df_{\mathrm{i}}^{\mathrm{eq}}(V)/dV]_0}{-i\omega + 1/\tau_{\mathrm{i}}(V_0)}\delta\tilde{V}(\omega) + \tilde{I}_{\mathrm{ext}}(\omega). \tag{514}$$

Collecting terms, we find

$$\left[-i\omega C + \frac{1}{R_0} + \sum_{\mathrm{i}} \frac{g_{\mathrm{i}} N_{\mathrm{i}}(V_0 - V_{\mathrm{i}})[df_{\mathrm{i}}^{\mathrm{eq}}(V)/dV]_0}{-i\omega + 1/\tau_{\mathrm{i}}(V_0)}\right]\delta\tilde{V}(\omega) = \tilde{I}_{\mathrm{ext}}(\omega). \tag{515}$$

The resting resistance of the membrane is defined by

$$\frac{1}{R_0} = \sum_{\mathrm{i}} g_{\mathrm{i}} N_{\mathrm{i}} f_{\mathrm{i}}^{\mathrm{eq}}(V). \tag{516}$$

The term in brackets in Eq (515) is the inverse impedance (or "admittance") of the system.

To understand what is going on here, it's useful to think about channels which have fast ($1/\tau_{\mathrm{i}} \gg \omega$) or slow ($1/\tau_{\mathrm{i}} \ll \omega$) responses. The fast channels renormalize the resistance,

$$\frac{1}{R_0} \to \frac{1}{R_0} + \sum_{\mathrm{i}\in\mathrm{fast}} \tau_{\mathrm{i}}(V_0) g_{\mathrm{i}} N_{\mathrm{i}}(V_0 - V_{\mathrm{i}}) \left.\frac{df_{\mathrm{i}}^{\mathrm{eq}}(V)}{dV}\right|_{V=V_0}. \tag{517}$$

Importantly, the correction to the resistance can be either positive or negative. Suppose that, as in Fig 98, the channels tend to open in response to increasing voltage, as most channels do. Then $[df_{\mathrm{i}}^{\mathrm{eq}}(V)/dV]_0 > 0$. But if this channel is specific for an ion with a reversal potential above the resting potential ($V_{\mathrm{i}} > V_0$), then opening the channel creates a stronger tendency to pull the voltage toward this higher potential, which is a regenerative effect—a negative resistance.

If the channels are slow, they make a contribution to the imaginary part of the admittance, along with the

capacitance,

$$-i\omega C \to -i\omega C + \frac{1}{-i\omega}\sum_{\mathrm{i}\in\mathrm{slow}} g_{\mathrm{i}}N_{\mathrm{i}}(V_0 - V_{\mathrm{i}})\frac{df_{\mathrm{i}}^{\mathrm{eq}}(V)}{dV}\bigg|_{V=V_0}. \tag{518}$$

Again the sign depends on details. If the channels are opened by increasing voltage and the reversal potential is *below* the resting potential, then their contribution is (almost) like an inductance, and can generate a resonance by competing with the capacitance. This resonance is at a frequency

$$\omega_* = \left[\frac{1}{C}\sum_{\mathrm{i}\in\mathrm{slow}} g_{\mathrm{i}}N_{\mathrm{i}}(V_0 - V_{\mathrm{i}})\frac{df_{\mathrm{i}}^{\mathrm{eq}}(V)}{dV}\bigg|_{V=V_0}\right]^{1/2} \tag{519}$$

which, interestingly, does not depend on the precise value of the time constants defining the channel kinetics, although one must obey the condition $\omega_* \gg 1/\tau_{\mathrm{i}}(V_0)$ for all $\mathrm{i} \in$ slow.

Problem 99: Equivalent circuits. Equation (515) shows that each type of channel contributes a parallel path for current flow through the membrane. The impedance of this path is defined by

$$\frac{1}{\tilde{Z}_{\mathrm{i}}(\omega)} = g_{\mathrm{i}}N_{\mathrm{i}}f_{\mathrm{i}}^{\mathrm{eq}}(V) + \frac{g_{\mathrm{i}}N_{\mathrm{i}}(V_0 - V_{\mathrm{i}})[df_{\mathrm{i}}^{\mathrm{eq}}(V)/dV]_0}{-i\omega + 1/\tau_{\mathrm{i}}(V_0)}. \tag{520}$$

Without resorting to the fast/slow approximations above, draw an equivalent circuit using the standard lumped elements (capacitance, resistance, inductance) which realizes this impedance. Show how the parameters of the lumped elements relate to the parameters of the channels.

So, we have seen that even in response to small signals, the dynamics of ion channels generate an interesting complement of electronic parts: resistors, inductors, and negative resistors. Certainly one can put these together to make a filter, playing the effective inductance of the channels against the intrinsic capacitance of the membrane, as noted above. With the negative resistor one can sharpen the resonance, and even generate an instability; presumably on the other side of the instability is a genuine oscillator.

Problem 100: Oscillations. Construct a minimal model for ion channels in the cell membrane that supports a stable, limit cycle oscillation of the voltage.

The negative resistance alone means that we can have (without oscillations) an instability of the steady state around which we were expanding, presumably because the real system is multi–stable. To see this more clearly, consider just two types of channels—a 'leak' channel which is open independent of the voltage and has a reversal potential of zero, and some other channel which opens in response to increasing voltage. Then the dynamics are

$$C\frac{dV}{dt} = -G_{\mathrm{leak}}V - gNf(V - V_{\mathrm{r}}), \tag{521}$$

$$\frac{df}{dt} = -\frac{1}{\tau(V)}[f - f_{\mathrm{eq}}(V)]. \tag{522}$$

The steady state solutions are determined by solving two simultaneous equations, usually called the nullclines, obtained by setting the time derivatives equal to zero:

$$f = f_{\mathrm{eq}}(V) \tag{523}$$

$$V = V_{\mathrm{r}}\frac{f}{f + G_{\mathrm{leak}}/gN}; \tag{524}$$

these are shown schematically in Fig 99, for some reasonable choice of parameters. Evidently there are three solutions to the two simultaneous equations, and it is fairly easy to show that two are stable and one is unstable. The two stable states correspond, roughly, to one state in which all the channels are closed and the voltage is zero (the reversal potential of the leak), and one state in which all the channels are open and the voltage is near the reversal potential for these channels. The bistability means that, if the cell starts in the low voltage state, injection of a relatively small, brief current can drive the system across a threshold (separatrix) so that it falls into the high voltage state after the current pulse is complete. This is a form of memory (interesting, although not very realistic), but also a substantial amplification of the incoming signal, especially if the parameters are tuned so that the difference in voltage to the unstable state is small.

Problem 101: Bistability. Work through a concrete example of the ideas in the previous paragraphs, perhaps using the detailed model from Fig 99. You should be able to verify, analytically, the claims about stability of the different steady states. Explain how these analytic criteria can be converted into a test for stability of each steady state that can be 'read off' directly from the plots in Fig 99. Analyze the response to brief pulses of current, showing that there is a well defined threshold for switching from one stable state to the other.

All the different kinds of dynamics we have seen thus far—filtering, oscillation, and bistability—can be generated by just one kind of channel with only two states.

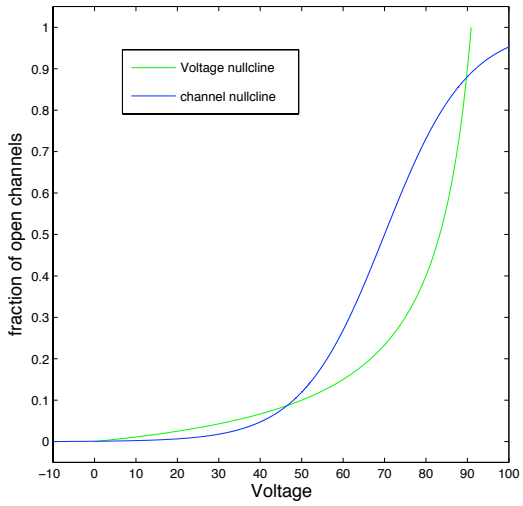

FIG. 99 Bistability in a simple model of a neuron. The channel nullcline is Eq (523), and the voltage nullcline is Eq (524). To be explicit we choose $f_{\rm eq}(V)$ from Eq (505), with $V_{1/2} = 70$ and $V_w = 10$, and $G_{\rm leak}/gN = 0.1$. Note that there are three crossing points, corresponding to steady states. The low voltage and high voltage states are stable; the intermediate voltage state is unstable.

Real neurons are much more complex. One important class of dynamics that we can't quite see in the simplest models is 'excitability.' In this case, a small pulse again drives the system across a threshold, but what would have been a second stable state is destabilized by relaxation of some other degrees of freedom; the result is that the system takes a long, and often stereotyped, trajectory through its phase space before coming back to its original steady state after the input pulse is over. The action potential is an example of such excitable dynamics [should we have a sketch of what this means in a simple phase plane?].

Our understanding of ion channels goes back to the classic work of Hodgkin and Huxley in the 1940s and 50s. They studied the giant axon, a single cell, visible to the naked eye, which runs along the length of a squid's body, and along which action potentials are propagated to trigger the squid's escape reflex. Passing a conducting wire through the interior of the long axon, they short–circuited the propagation, insuring that the voltage across the membrane was spatially uniform, as in our idealization above. They then studied the current that flowed in response to steps of voltage. If the picture of channels is correct, then with the voltage held constant, there should be an (Ohmic) flow of current through the open channels. If we step suddenly to a new value of the voltage, Ohm's law tell us that the current through the open channels will change immediately, but there will be a prolonged time dependence that results from the open or closing of channels as they equilibrate at the new voltage. In the simple model with two states, this changing current should relax exponentially to a new steady state; in particular, the initial slope of the current should be finite.

Hodgkin and Huxley found that the relaxation of the current at constant voltage has a gradual start, as if the channels had not one closed state but several, and the molecules had to go through these states in sequence before opening. They chose to describe these dynamics of the currents by imagining that, in order for the channel to be open, there were several independent molecular "gates" that all had to be open. Each gate could have only two states, and would obey simple first order kinetics, but the probability that the channel is open would be the product of the probabilities that the gates were open. In the simple case that the multiple gates are identical, the probability of the channel being open is just a power of the 'gating variable' describing the probability that one gate is open. Hodgkin and Huxley also discovered that at least one important class of channels open in response to increased voltage, and then seem to close over time. They described this by saying that in addition to 'activation gates' that were opened by increasing voltage, there were 'inactivation gates' which closed in response to increasing voltage, but these had slower kinetics. Putting the pieces together, they described the fraction of open channels as

$$f_{\rm i} = m_{\rm i}^{\alpha_{\rm i}} h_{\rm i}^{\beta_{\rm i}}, \tag{525}$$

where m and h are activation and inactivation gates, respectively, and the powers α and β count the number of these gates that contribute to the opening of one channel. The kinetics are then described by

$$\frac{dm_{\rm i}}{dt} = -\frac{1}{\tau_{\rm i}^{(m)}(V)}\left[m_{\rm i} - m_{\rm i}^{\rm eq}(V)\right] \tag{526}$$

$$\frac{dh_{\rm i}}{dt} = -\frac{1}{\tau_{\rm i}^{(h)}(V)}\left[h_{\rm i} - h_{\rm i}^{\rm eq}(V)\right], \tag{527}$$

and finally the voltage (again neglecting spatial variations) obeys

$$C\frac{dV}{dt} = -\sum_{\rm i} g_{\rm i} N_{\rm i} m_{\rm i}^{\alpha_{\rm i}} h_{\rm i}^{\beta_{\rm i}} (V - V_{\rm i}). \tag{528}$$

Problem 102: Two gates. Suppose that each channel has two independent structural elements ("gates"), each of which has two states. Assuming that the two gates are independent of one another, fill in the steps showing that the dynamics of the channels are as described above. In particular, show that after a sudden change in voltage, the fraction of open channels starts to change as $\propto t^2$, not $\propto t$ as expected if the entire channel only has two states. [This, and the preceding paragraph, might be a little too telegraphic. Need feedback here!]

Problem 103: Hodgkin and Huxley revisited. The original equations written by Hodgkin and Huxley are as follows:[69]

$$C\frac{dV}{dt} = -\bar{g}_L(V - V_L) - \bar{g}_{\rm Na}m^3h(V - V_{\rm Na}) - \bar{g}_{\rm K}n^4(V - V_{\rm K}) + I(t) \quad (529)$$

$$\frac{dn}{dt} = (0.01V + 0.1)\exp(-V/10)(1 - n) - 0.125n\exp(V/80)n \quad (530)$$

$$\frac{dm}{dt} = (0.1V + 2.5)\exp(-V/10 - 1.5)(1 - m) - 4\exp(V/18)m \quad (531)$$

$$\frac{dh}{dt} = 0.07\exp(V/20)(1 - h) - \exp(-V/10 - 4)h, \quad (532)$$

where Na and K refer to sodium and potassium channels, respectively; time is measured in milliseconds and V is measured in milliVolts. These equations are intended to describe a small patch of the membrane, and so many parameters are given per unit area: $C = 1\,\mu{\rm F/cm^2}$, $\bar{g}_L = 0.3\,{\rm mS/cm^2}$, $\bar{g}_{\rm Na} = 120\,{\rm mS/cm^2}$, and $\bar{g}_{\rm K} = 36\,{\rm mS/cm^2}$; the reversal potentials are $V_L = 10.613\,{\rm mV}$, $V_{\rm Na} = 115\,{\rm mV}$, and $V_{\rm K} = -12\,{\rm mV}$.

(a.) Rewrite these equations in terms of equilibrium values and relaxation times for the gating variables, e.g.

$$\frac{dm}{dt} = -\frac{1}{\tau_m(V)}\left[m - m_{\rm eq}(V)\right]. \quad (533)$$

Plot these quantities. Can you explain, intuitively, the form of the curves?

(b.) Simulate the dynamics of the Hodgkin–Huxley equations in response to constant current inputs. Show that there is a threshold current, above which the system generates period pulses. Explore the frequency of the pulses as a function of current.

(c.) Suppose that the injected current consists of a mean (less than the threshold you identified in [b]), plus a small component at frequency ω. By some appropriate combination of analytic and numerical methods, find the impedance $Z(\omega)$ for different values of the mean injected current. Show that the membrane has a resonance, and explore what happens to this resonance as the mean current is increased toward threshold. How do your results connect to the frequency of pulses above threshold?

(d.) Real axons are essentially long thin cylinders. Show that, if we allow the voltage to vary along the length of the axon, there should be a current per unit area flowing across the membrane of

$$I = \frac{a}{2R}\frac{\partial^2 V}{\partial z^2}, \quad (534)$$

where z is the coordinate along the cylinder, a is its radius, and R is the resistivity of the fluid filling the axon, assuming that resistance outside the axon is negligible. For the squid giant axon, $a \sim 250\,\mu$m and $R \sim 35\,\Omega\cdot$cm. Use this result to write equations for the voltage and gating variables along the axon. Note that only the dynamics of voltage is sensitive to spatial derivatives. Why?

(e.) Simulate the response of a long segment of the axon to a current pulse injected at one end. Show that small pulses result in spatially restricted voltage responses, while larger pulses produce a propagating pulse. Confirm that these pulses become more stereotyped as the propagate, and have a velocity that is independent of the input current. What is this velocity? How does it compare to the observed speed of action potentials, $v \sim 20\,{\rm m/s}$?

[69] The only difference from the original paper is that we use the modern sign convention for the voltage. Notice that this original formulation is in terms of a "maximal conductance" for each type of "current," while in modern language we could talk about the number of each type of channel. In fact, the more phenomenological description persists, because it corresponds more directly to what is measured, but this allows us to forget that parameters such as $\bar{g}_{\rm K}$ actually measure the number of copies of a protein that have been inserted into the membrane.

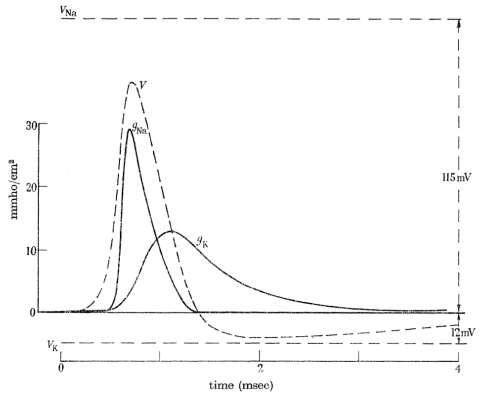

FIG. 100 The action potential that emerges from the Hodgkin–Huxley model. Need to decide what to say, what other things to reproduce ... .

Problem 104: Simplification. It is very hard to make analytic progress in understanding the dynamics of a system with five variables. There is a history of trying to approximate the system by exploiting the fact that the different variables have very different time scales. See how far you can go along this path. I have left this problem deliberately open–ended. For one approach, see Abbott and Kepler (1990).

It is good to pause here and review how we know that the Hodgkin–Huxley description of ion channels is correct. [Not sure how much of this should be illustrated by figures from the original papers?] The initial triumph, which you are asked to reproduce in the problem above, is the prediction of the propagating action potential itself, as in Fig 100, with the correct speed. The model also predicts that, as the action potential passes, there is a net flux of potassium and sodium across the membrane. On long time scales, this must be balanced by the action of pumps that maintain the concentration differences between the inside and outside of the cell. But either by looking quickly or by poisoning the pumps, one should be able to detect the flux, for example using radioactive tracers, and this works, quantitatively.

[This is all a little vague; should go back and try to do better!] Nature provides a variety of toxins which block the action potential in different ways, and we can also find artificial blockers, for example using ions with very large radius that can literally plug the hole in open channels. It is striking that these agents act selectively on different channels, and one can verify that this way of isolating the dynamics of sodium and potassium channels matches the Hodgkin–Huxley description. If we can arrange for the channels to "open" but be blocked, then the structural change of the channel molecule upon opening should still

move the gating charge across the membrane, and if we are careful this should be measurable essentially as a delayed capacitive response to changes in the applied voltage. These "gating currents" have indeed been detected, and in some cases it has been possible to match these quantitatively not only against predictions based on the form of the activation curve, but also to genetically engineer the channels and show that changes in the activation curve and gating currents track one another. [How much detail here? Give the example of shaker?]

If individual channels are independent of one another, then their opening and closing events should be independent. If we look at a small patch of the membrane, there will not be that many channels present, and we might be able to see that the discrete events in the individual molecules don't quite average out—there should be noise from the random opening and closing of the single channels. This channel noise has been detected, and has the spectral properties expected from the Hodgkin–Huxley model. Finally, if we look at even smaller patches of the membrane, and have proportionately more sensitive amplifiers, we should be able to see the opening and closing of single channels. Again, this works. Most importantly, we can look at the distribution of times that individual channels spend in the open and closed states, and connect this to the kinetics predicted by the Hodgkin–Huxley model and its generalizations. Although these more detailed measurements have revealed new features of channel kinetics even in well studied examples, in outline the picture given to us by Hodgkin and Huxley has stood the test of time. [Again, should probably show some figures. Emphasize how remarkable it is to be looking at individual molecular events—current flow through sub–nanometer pores! Maybe even discuss shot noise through open channels?]

Problem 105: Channel noise. Give a problem that maps the HH model onto a stochastic picture of channel states, and then derive the expected properties of the channel noise. Remember that we did the simplest version of this in Chapter 1.

Problem 106: Single channel kinetics. Give a problem that explores how single channel kinetics are connected to the macroscopic kinetics.

Now that we have confidence in our mathematical description of neurons, it is time to realize now just how many parameters are involved. A typical cell expresses eight or nine different kinds of channels. Each channel is described by the dynamics of two gating variables. If we imagine that activation or inactivation curves have the simple sigmoidal form as in Fig 99, then there are roughly two parameters for each such curve—the voltage at half activation and the slope or width—and at least one more parameter to set the time scale of the kinetics. Finally, there is the total number of channels, or the maximum conductance achieved if all the channels are open. All together, then, this is ~ 7 parameters per channel type, or roughly fifty parameters for the entire neuron, conservatively. Importantly, to a large extent the cell actually has control over these parameters, and, in a meaningful sense, can adjust them almost continuously.

How do these adjustments occur? Most obviously, the total number of open channels is controlled in the same that all other protein copy numbers are controlled. Sometimes, because of the clearer connection to experiment, one speaks about the 'maximal conductance' associated with a particular type of channel ($G_i^{max} = g_i N_i$), but this obscures the fact that this parameter really is the total number of copies of the protein that the cell has expressed and inserted into the membrane. The parameters of the activation curves and the time constants are intrinsic properties of the proteins, but these too can be adjusted in several ways. First, like all proteins, ion channels can be covalently modified by phosphorylation etc.. More importantly, the genome encodes a huge number of different ion channels proteins; the human genome has 90 different potassium channels alone. While these do form classes based on their dynamics, there is considerable variation within classes, and since many of these genes have multiple alternative splicings, there is the potential for almost continuous parameter variation. These different mechanisms of variation interact; as an example, different splicing variants can exhibit different sensitivity to phosphorylation.

Problem 107: Continuous adjustment of electrical dynamics. [It might be that I should take the students by hand through the model; let's see how this works.] To illustrate the possibility of nearly continuous adjustments in the electrical dynamics of neurons, consider the case of the hair cells in the turtle ear. In these cells (cf Section 2.5), one contribution to frequency selectivity comes from a resonance in the electrical response of the hair cell itself. This resonance is driven by a combination of voltage–gated calcium channels and calcium–activated potassium channels. There is a detailed model of this system, described by Wu & Fettiplace (2001). Try to understand what they have done, and reproduce the essential theoretical results. In particular, what is the role of "details" (e.g., the building of channels out of combinations of different subunits) in generating the correct qualitative behavior?

One well studied example of channel dynamics is in the stomatogastric ganglion of crabs and lobsters, schematized in Fig 101. This is a network of ~ 30 neurons which generates a rhythm, and this rhythm in turn drives muscles which actuate teeth in the crab stomach, grinding its

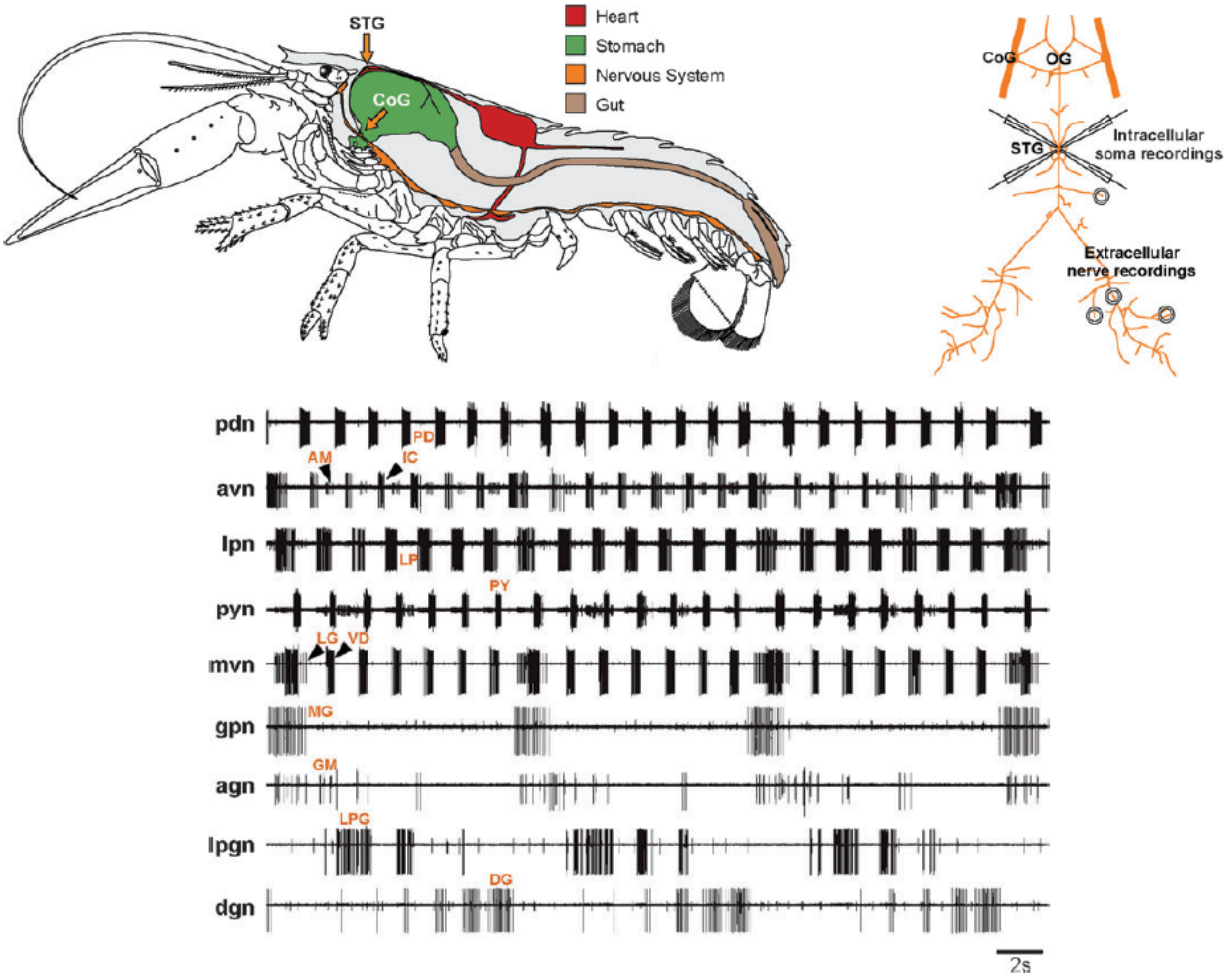

FIG. 101 The stomatogastric ganglion (STG) in crustaceans, from Marder & Bucher (2007). At top left, the location of the STG and the commissural ganglion (CoG) in a lobster. A top right, a schematic of the ganglion dissected out of the animal, and the opportunities for recording the activity of the neurons. At bottom, simultaneous extracellular recordings from nine motor nerves at the output of this network. Names indicate particular neurons which can be identified in each individual (as with the named neurons in the fly visual system discussed in [pointer]), and in some cases (e.g., avn, mvn) we can identify spikes from several individual neurons in the recording from one nerve. There are two main rhythms, the faster pyloric rhythm in cells PD, LP, PY, VD and IC, and the slower gastric mill rhythm in cells MG, DG, GM, LPG and LG.

food. Evidently getting the correct rhythm is important in the life of the organism. If one records the electrical signals from individual neurons, several of the cells produce period bursts of action potentials, and a handful of cells are 'pacemakers' that can generate this periodic pattern without input from the other cells. In one such cell (the lateral pyloric neuron), experiments show that there are seven different channel types. An important feature of this cell, shared by many other cells, is the presence of voltage–gated calcium channels. This means that, as action potentials occur, they trigger calcium flux into the cell. Because there are also channels which are directly affected by the calcium concentration, a complete model must include a description of the calcium buffering or pumping that counterbalances this flux.

It is worth being very explicit about all these ingredients in the dynamics of the lateral pyloric neuron, not least to get a sense for the state of the art in such analyses. As before, we will neglect the spatial structure of the cell, so there is just one relevant voltage difference V between the inside and outside of the cell, which obeys a slight generalization of Eq (528),

$$C\frac{dV}{dt} = -\sum_{\rm i} g_{\rm i} N_{\rm i} m_{\rm i}^{\alpha_{\rm i}} h_{\rm i}^{\beta_{\rm i}} (V - E_{\rm i}) + I_{\rm ext}, \qquad (535)$$

where $I_{\rm ext}$ is any externally injected current and $E_{\rm i}$ is the reversal potential for channel type i. The kinetics of the gating variables $m_{\rm i}$ and $h_{\rm i}$ are governed by Eq's (526) and (527), respectively. For most of the channels, we can take the equilibrium values of the gating variables to be

channel type	$g_i N_i$ (μS)	E_i (mV)	midpoints (mV)	widths (mV)	rates (s^{-1})
i = 1: **"delayed rectifier"**	0.35	$E_K = -80$			
activation equilibrium ($\alpha_1 = 4$)			$V_{1/2}^{m_1} = -25$	$V_w^{m_1} = 17$	
activation kinetics			$V_1^{(m)} = 10$	$1/\gamma_1^{(m)} = 22$	$k_1^{(m)} = 180$
i = 2: **Ca^{++} current 1**	0.21	E_{Ca}			
activation ($\alpha_2 = 1$)			$V_{1/2}^{m_2} = -11$	$V_w^{m_2} = 7$	50
inactivation ($\beta_2 = 1$)			$V_{1/2}^{h_2} = -50$	$V_w^{h_2} = -8$	16
i = 3: **Ca^{++} current 2**	0.047	E_{Ca}			
activation ($\alpha_3 = 1$)			$V_{1/2}^{m_3} = -22$	$V_w^{m_3} = 7$	10
i = 4: **"inward rectifier"**	0.037	-10			
activation equilibrium ($\alpha_4 = 1$)			$V_{1/2}^{m_4} = -70$	$V_w^{m_4} = -7$	
activation kinetics			$V_4^{(m)} = -110$	$1/\gamma_1^{(m)} = 13$	$k_1^{(m)} = 0.33$
i = 5: **"leak"**	0.1	-50			
i = 6: **"A–current"**	2.2	$E_K = -80$			
activation equilibrium ($\alpha_6 = 3$)			$V_{1/2}^{m_6} = -12$	$V_w^{m_6} = 26$	
activation kinetics					$k_6^{(m)} = 140$
inactivation equilibrium ($\beta_{6a} = 1$)			$V_{1/2}^{h_{6a}} = ...$	$V_w^{m_{6a}} = ...$	
inactivation kinetics					$k_{6a}^{(h)} = ...$
inactivation equilibrium ($\beta_{6b} = 1$)			$V_{1/2}^{h_{6b}} = ...$	$V_w^{m_{6b}} = ...$	
inactivation kinetics					$k_{6b}^{(h)} = ...$

TABLE I A subset of channels in the lateral pyloric neuron, from Buchholtz et al (1992). For the delayed rectifier and the second type of calcium channel, there is no evidence for inactivation. The negative value of $V_w^{(h_2)}$ means, from Eq (505), that the probability of the inactivation gate being "open" decreases with increasing voltage. For calcium channels, the reversal potential varies, depending on the calcium concentration inside the cell, as in Eq (541), and the relaxation times do not have a detectable voltage dependence. The voltage dependence of the inward rectifier kinetics is opposite to Eq (538), that is $1/\tau \propto 1 + \exp[-\gamma_i^{(m)}(V - V_i^{(m)})]$. The leak current, by convention, is the current that exhibits no voltage or time dependence of its conductance. Get details of the A–current right!

given by the generalization of Eq (505),

$$m_i^{eq}(V) = \frac{1}{1 + \exp[-(V - V_{1/2}^{m_i})/V_w^{m_i}]}, \quad (536)$$

$$h_i^{eq}(V) = \frac{1}{1 + \exp[-(V - V_{1/2}^{h_i})/V_w^{h_i}]}, \quad (537)$$

and the time constants for relaxation of the gating variables are, phenomenologically,

$$\frac{1}{\tau_i^{(m)}(V)} = \frac{k_i^{(m)}}{1 + \exp[-\gamma_i^{(m)}(V - V_i^{(m)})]}, \quad (538)$$

$$\frac{1}{\tau_i^{(h)}(V)} = \frac{k_i^{(h)}}{1 + \exp[-\gamma_i^{(h)}(V - V_i^{(h)})]}. \quad (539)$$

As shown in Table I, this description works for several channel types, one selective for potassium, two for calcium, and one mixed, plus a "leak" that exhibits no significant time or voltage dependence of its conductance.

Two of the important channel types allow calcium to flow into the cell. As we will see, this current is big enough to change the concentration of calcium inside the cell, and this has a variety of effects on other processes, including one of the channels that doesn't fit the simple description we have given so far. So, we will need to describe the dynamics of the calcium concentration itself. The simplest model is that the calcium relaxes back to some internally determined steady state, $[Ca]_0 = 0.05\,\mu M$, with a rate $k_{Ca} = 360\,s^{-1}$, in which case

$$\frac{d[Ca]}{dt} = -k_{Ca}([Ca] - [Ca]_0) + AI_{Ca}, \quad (540)$$

where I_{Ca} is the total calcium current ($I_{Ca} = I_2 + I_3$ from Table I). The constant $A = 300\,\mu M/nC$ is inversely proportional to the volume into which the current flows, which experimentally comes out to be much smaller than the total volume of the cell body. As the concentration of calcium changes, the reversal potential for the calcium currents also changes,

$$E_{Ca} = \frac{k_B T}{2e} \ln\left(\frac{[Ca]_{out}}{[Ca]}\right), \quad (541)$$

where the calcium concentration outside the cell is $[Ca]_{out} = 13\,mM$.

We are still missing three of the channel types in this cell. First, there is another potassium channel that is almost described by our standard model, but the inactivation seems to involve two processes that occur on different time scales. This can be captured by replacing

$$h_6 \to x(V)h_{6a} + [1 - x(V)]h_{6b}, \tag{542}$$

where the weighting function

$$x(V) = \frac{1}{1 + \exp[-(V - 7)/15]}, \tag{543}$$

with V measured in mV as before.

Next, there is a fast sodium channel not unlike the ones that Hodgkin and Huxley found in the squid giant axon, with $\alpha_7 = 3$ and $\beta_7 = 1$. The activation is sufficiently fast that it can be approximated as instantaneous, so that m_7 is always at its equilibrium value, which varies with voltage in a slightly more complicated way than for the other channels,

$$m_7 = m_7^{\rm eq}(V) = \frac{1}{1 + \frac{136}{V+6}\left(\exp[-(V + 34)/13] - \exp[-(V - 0.07)/7.9]\right)}, \tag{544}$$

where V again is measured in mV [Need to check this carefully!]. The inactivation gates obey

$$\frac{dh_7}{dt} = a_7(V)(1 - h_7) - b_7(V)h_7, \tag{545}$$

where the rates

$$a_7(V) = 40\exp[-(V + 39)/8], \text{ and} \tag{546}$$

$$b_7(V) = \frac{500}{1 + \exp[-(V + 40)/5]}, \tag{547}$$

are measured in s^{-1}. The total conductance that is contributed by these channels is large, $g_7 N_7 = 2300\,\mu$S, although they are only open briefly.

The last type of channel, like the first two in Table I, is selective for potassium ions, but the probability of the channel being open is modulated by the intracellular calcium concentration. This channel has $\alpha_8 = \beta_8$, and the equilibrium state of the inactivation gate depends only on the calcium concentration,

$$h_8^{\rm eq} = \frac{1}{1 + [\mathrm{Ca}]/(0.6\,\mu\mathrm{M})}. \tag{548}$$

The equilibrium state of the activation gate, in contrast, depends both on voltage and on calcium,

$$m_8^{\rm eq} = \frac{1}{1 + \exp[-(V + f[\mathrm{Ca}])/23]} \cdot \frac{1}{1 + \exp[-(V + 16 + f[\mathrm{Ca}])/5]} \cdot \frac{[\mathrm{Ca}]}{2.5\,\mu\mathrm{M} + [\mathrm{Ca}]}, \tag{549}$$

where $f = 0.6\,\mathrm{mV}/\mu\mathrm{M}$. The relaxation rates $k_8^{(m)} = 600\,\mathrm{s}^{-1}$ and $k_8^{(h)} = 35\,\mathrm{s}^{-1}$ show little if any voltage dependence. This seems like a complicated model, but it fits the experimental results very well, as in Fig 102.

Problem 108: Calcium dependent potassium conductances. Develop a microscopic picture to explain the combination of voltage and calcium dependences seen in Eq's (548) and (549). Remember that these equations describe the equilibrium fractions of molecules in particular states, so you need to relate these back to the free energies of the different states. Connect your discussion with the MWC models discussed in Appendix A.4 and [elsewhere?].

The model of the lateral pyloric neuron which we have described here represent the culmination of many years of effort, both in experiments on this particular system and in the exploration of these fully realistic generalizations of the Hodgkin–Huxley model to what seems the more typical case, with many different channel types functioning together. This model also represents a level of detail and complexity that I have tried to avoid so far, so some explanation is called for. First, the complexity consists largely of variations on a theme. Many channels are known to be described by the general picture of multiple activation and inactivation gates, so this provides a framework within which each new type of channel can be fit. Second, the complexity is justified by a large body of data. There are independent experiments on other systems, exploring quantitatively each of the types of channels that we see in this neuron, and detailed experiments

on this one cell to tease out the contributions of each of the channel types.

Problem 109: Justifying complexity. Go through Golowasch & Marder (1992), Buchholtz et al (1992), and Golowasch et al (1992), and explain the justification for each of the channel types in the model discussed above.

Indeed, the program of describing the electrical dynamics of single neurons in terms of generalized Hodgkin–Huxley models, usually with many different channel types functioning together, became a small industry. It really worked. In some cases one could go so far as to characterize the kinetics of particular channel types through measurements on single molecules, and then put these single molecule properties together to reproduce the functional behavior of the cell as a whole. This really is quite a beautiful body of work, and implements what many people would like to do in other systems, building from measured properties of individual molecular events up to macroscopic biological function. As emphasized above, we can think of the ion channels in the cell membrane as a network of interacting proteins, where the interaction is mediated by the voltage across the membrane rather than direct protein–protein encounters, and where the equations for the dynamics of the individual channels have a firm foundation. It is not unreasonable to claim that ion channels in the cell membrane are in fact the best understood examples of biochemical networks, although the language typically used in describing these systems obscures this connection.

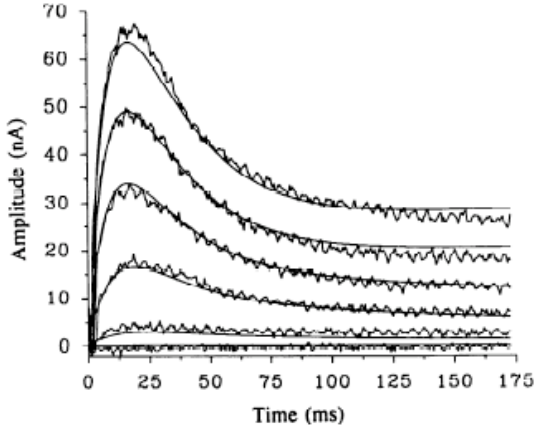

FIG. 102 Dynamics of the calcium dependent potassium current, from Buchholtz et al (1992). Experimental data (noisy traces) from Golowasch & Marder (1992), solid lines from the model including Eq's (548) and (549). [Go back and understand how they isolate this contribution to the current]

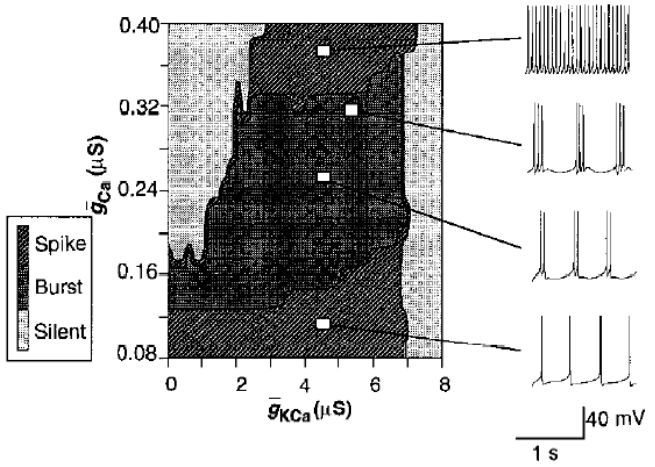

FIG. 103 Simulations of a detailed model, with seven types of channel, for the lateral pyloric neuron in the stomatogastric ganglion of the crab. Changes in the pattern of activity as a function of the numbers of two different kinds of channel, where channel number here is expressed as the maximal conductance when all channels are open. Note that relatively small changes in these parameters can result in both quantitative and qualitative changes in the pattern of electrical activity, running the full range from silence to single spike firing to bursting. From Le Masson et al (1993).

Despite their success, it came to be known, though not widely commented upon, that these models of coupled ion channel dynamics had a problem. While experiments often characterize the activation curves and kinetics of the individual channels, it is hard to make independent measurements of the total number of channels, or equivalently the maximum conductance when all the channels are open. Thus, one is left adjusting these parameters, trying to fit the overall electrical dynamics of the neuron—for example, the rhythmic bursting of the pyloric neuron. This fitting turns out to be delicate; as one adjusts the (many) parameters, one finds bifurcations to qualitatively different behaviors in response to relatively small changes. An example of this is shown in one two–dimensional slice through the seven dimensional space of channel numbers in the pyloric model, at the top in Fig 103.

Frankly, from a physicist's point of view this all seems a mess. There are many details one has to keep track of, and many parameters to adjust. One might be tempted just to walk away, and count this as a part of biology we don't want to know about. But there is a deep ques-

tion here:[70] if we have trouble adjusting the parameters of our models in order to reproduce the observed functional behaviors of particular cells, how do the cells themselves adjust these parameters to achieve their correct functions? How does it choose the 'correct' number of each type of channel to express? One could imagine that the cell has some sort of lookup table—I am a cell of type α, so I should express N_1^α molecules of channel type 1, N_{37}^α molecules of channel type 37, and so on. This is a bit implausible. More likely would be that the cell has a way of monitoring its activity and asking "how close am I to doing the right thing?," generating an error signal that could be used to drive changes in the expression of the channels or perhaps their insertion into the membrane.

How can a neuron "know" whether it is exhibiting the desired pattern of electrical activity? It would need some signal that couples voltage changes across the membrane, which are quite fast, to the biochemical events regulating gene expression, which are quite slow. One idea is to use the intracellular calcium concentration as an intermediary. We know that many cellular processes are regulated by calcium, so one end of this is easy to imagine. But in the models described above the calcium concentration is an explicit part of the dynamics, so we can calculate, for example, the time average calcium concentration as function of the parameters of the model. What we see in Fig 104 is that $[Ca^{++}]$ does an excellent job of tracing the pattern of electrical activity in this cell. Thus if the system wants to stabilize a pattern of rhythmic bursting, it can do so via feedback mechanisms which try to hold the calcium concentration near a target value of $C_0 \sim 0.2\,\mu$M.

Let us suppose that the expression of each channel protein is regulated by calcium, so that

$$\tau_i \frac{dN_i}{dt} = N_i^{max} f_i([Ca^{++}]/C_0) - N_i, \qquad (550)$$

where $f_i(x)$ is a sigmoidal function such as

$$f_i(x) = \frac{1}{1 + x^{\pm n}}. \qquad (551)$$

Of course these equations have their steady state at $N_i = N_i^{max} f_i([Ca^{++}]/C_0)$, but the calcium concentration must be determined self–consistently through the full dynamics of the channels and voltage. We should choose the signs of the calcium dependences to insure stability: channels which allow excitatory currents to flow will tend to drive increases in $[Ca^{++}]$, and so these should be opposed by a decreasing function $f_i(x)$, and vice versa. Once we do this, if the regulation functions are steep [large value of n in Eq (551)], and the maximum possible numbers of channels (N_i^{max}) are large, the dynamics will always be pulled into regimes where $[Ca^{++}] \approx C_0$. We need a figure which illustrates this!

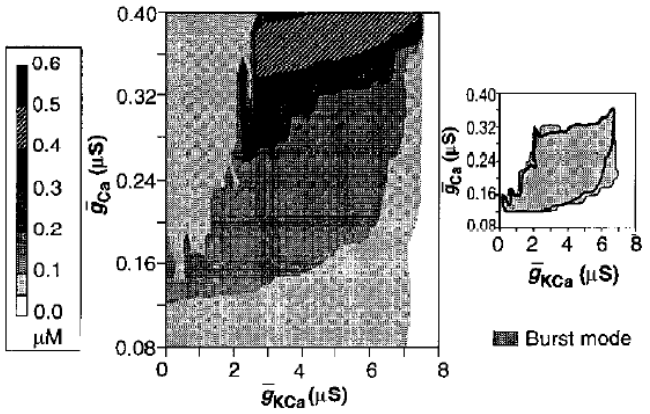

FIG. 104 Mean calcium concentration follows the pattern of electrical activity. Main figure shows the mean calcium concentration as a function of the same two variables shown in Fig 103. Small figure at right shows that the region of bursting activity corresponds almost perfectly to the region of parameter space in which the mean calcium concentration is between 0.1 and 0.3 μM, so that holding the calcium level fixed will stabilize bursting. From Le Masson et al (1993).

Problem 110: A simple example of a self–tuning neuron. Need to find the simplest example of these models, and let the students work it through for themselves.

How can we tell if something like this sort of self–tuning really is happening? If neurons knew how many of each kind of channel to make, then they would try to do this no matter what the conditions. For example, inputs from other neurons would drive changes in the electrical activity, but not changes in channel expression. On the other hand, if the cell is 'trying' to maintain some mean calcium concentration, or some other measure of activity, then changing the environment in which the neuron operates will change channel expression. As an extreme example, if we rip the neuron from its network and put it in a dish, the normal pattern of rhythmic bursting will go (wildly) wrong, but the calcium–sensitive dynamics of the channel expression levels will eventually bring the system back into something close to the original pattern.

[70] As in the case of kinetic proofreading, I think there is a tendency to remember the original papers as having proposed mechanisms that solve problems. But I think that, in many ways, it was a much deeper contribution to *formulate* the problems. Even if the solutions turn out not to be precisely the ones chosen by Nature, the problems are important.

In this new state, the channels are playing different roles in the dynamics, because the driving forces for ionic current flow are different, but the final pattern of activity is the same. A literal version of this rather dramatic scenario actually works experimentally, as shown in Fig 105.

We have noted already that, in invertebrates such as flies and crabs, neurons have names, numbers and identifiable functions from individual to individual within a species. This discussion of stabilizing patterns of activity rather than expression levels suggests that this reproducibility of function can be achieved without exactly reproducing the number of copies of each channel protein. Further, although the slice through parameter space shown in Fig 104 suggests that the region compatible with normal function is convex, this in fact is not the generic case, and real models often have banana–shaped volumes in parameter space which are consistent with particular patterns of electrical activity. [Look through Goldman et al (2001) & Golowasch et al (2002) to decide on a figure.] Again this is consistent with what one sees experimentally, most impressively in subsequent experiments which measure directly the number of copies of mRNA for several channel types in single cells [recent refs from Eve's group].

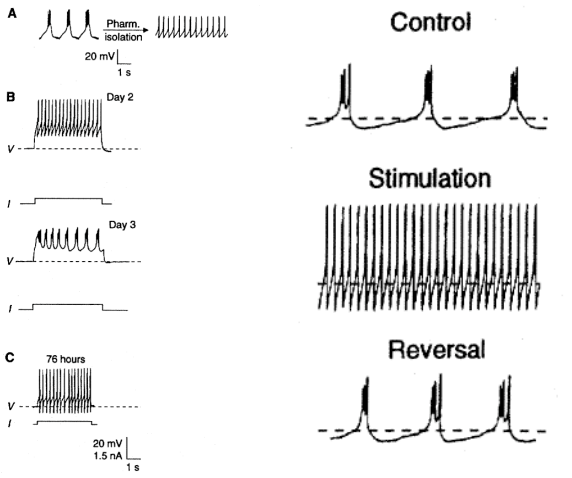

FIG. 105 Changing intrinsic properties of the STG neurons, from Turrigiano et al (1994). At left, an experiment in which one cell is ripped from the network and placed in isolation. At first (top) the electrical activity shifts from rhythmic bursts to repeated ("tonic") firing of single action potentials. After two days in culture, the cell is silent but responds to small positive currents with tonic firing; after three days the response consists of bursts not unlike those in the native network environment. At bottom, continuous recordings demonstrate that this switch from tonic firing to bursting can occur within an hour. At right, one hour of stimulation with negative current pulses drives a shift from bursting to tonic firing, which is reversed after one hour of no stimulation. All these changes in activity reflect changes in the numbers of different types of ion channels in the cell membrane, as predicted from the models discussed in the text.

One might worry that we have replaced the tuning of channel copy numbers with a fine tuning of the regulatory mechanisms on all the channels. In fact, it is not plausible that calcium acts directly on expression of genes. More likely is that calcium binds to some protein, and when its binding sites are occupied the protein can act, directly or indirectly, as a transcription factor. Then the fact that all the genes have the same calcium dependence to their steady state values reflects the fact that they are all being regulated by the same calcium binding protein. Exploring this scenario in more detail, one realizes that the kinetics of binding and unbinding of calcium to the sensitive protein can span the time scales of action potentials, bursts, and even the basic rhythm itself. By combining signals from calcium binding proteins with different kinetics [that's a little quick!] one can thus stabilize more subtle details in the pattern of electrical activity. Maybe there is more to say about all this before drawing the lessons. Check most recent papers.

Faced with a model that explains the behavior of cells only when parameters are finely tuned, we become suspicious that we are missing something. One possibility—often the most plausible—is that the model simply is wrong. The models that we have for biological systems are not like the Navier–Stokes equations for fluids or the standard model of particle physics; we have many reasons to doubt that we are simply solving the wrong equations. But the electrical dynamics of neurons are a special case. Our mathematical models of channel dynamics emerged as accurate summaries of a huge body of data, and are nearly exact on the time scales that are experimentally accessible. Rather than rejecting the models, we must conclude that we are missing something, presumably on time scales longer than the experiments that go into characterizing the channel kinetics. In particular, what look like constant parameters must become slow dynamical variables. The simplest implementation of this idea seems to work, and to generate several dramatic experimental predictions which have since been confirmed. Indeed, this theoretical work on the problem of parameter determination has launched a whole subfield of experimental neurobiology, investigating the activity–dependent regulation of the 'intrinsic' electrical properties of neurons [be sure there is a ref to recent review].

Our understanding of ion channels goes back to the classic papers of Hodgkin and Huxley (1952a–d), still very much worth reading. The series of papers (of which the first really is Hodgkin, Huxely & Katz 1952) describes many ingenious experiments, culminating in a mathematical model which predicts the form and speed of the action potential. Iinclude Hodgkin's summaries—Croonian lecture, plus the one from *Pursuit of Nature*] For a modern textbook account, see Dayan & Abbott (2001). The Hodgkin–Huxley model is

complicated, so over the years there have been various attempts at simplifying to the point where one can gain analytic insight; for one approach, see Abbott & Kepler (1990) [also FitzHugh & Nagumo].

Abbott & Kepler 1990: Model neurons: From Hodgkin–Huxley to Hopfield. LF Abbott & T Kepler, in *Statistical Mechanics of Neural Networks*, L Garrido, ed, pp 5–18 (Springer–Verlag, Berlin, 1990).

Dayan & Abbott 2001: *Theoretical Neuroscience* P Dayab & LF Abbott (MIT Press, Cambridge, 2001).

Hodgkin et al 1952: Measurement of the current–voltage relations in the membrane of the giant axon of *Loligo*. AL Hodgkin, AF Huxley & B Katz, *J Physiol (Lond)* **117,** 442–448 (1952).

Hodgkin & Huxley 1952a: Currents carried by sodium and potassium ions through the membrane of the giant axon of *Loligo*. AL Hodgkin & AF Huxley, *J Physiol (Lond)* **117,** 449–472 (1952).

Hodgkin & Huxley 1952b: The components of membrane conductance in the giant axon of *Loligo*. AL Hodgkin & AF Huxley, *J Physiol (Lond)* **117,** 473–496 (1952).

Hodgkin & Huxley 1952c: The dual effect of membrane potential on sodium conductance in the giant axon of *Loligo*. AL Hodgkin & AF Huxley, *J Physiol (Lond)* **117,** 497–506 (1952).

Hodgkin & Huxley 1952d: A quantitative description of membrane current and its application to conduction and excitation in nerve. AL Hodgkin & AF Huxley, *J Physiol (Lond)* **117,** 500–544 (1952).

For a modern view of ion channels, see Hille (2001). [add some classic references about resonances etc.] For a detailed discussion of system in which the effective resonance generated by channel kinetics has functional importance, see Wu & Fettiplace (2001). [Need references that survey the richness of ion channel diversity, phosphorylation, splicing variants, etc. Check Laughlin refs re splicing variants in the fly eye.] For one example of this complexity, see Tian et al (2001).

Hille 2001: *Ion Channels of Excitable Membranes, Third Edition*. B Hille (Sinauer, 2001).

Wu & Fettiplace 2001: A developmental model for generating frequency maps in the reptilian and avian cochleas. YC Wu & R Fettplace, *Biophys J* **70,** 2557–2570 (1996).

Tian et al 2001: Altenative splicing switches potassium channel sensitivity to protein phosphorylation. L Tian, RR Duncan, MS Hammon, LS Coghill, H Wen, R Rusinova, AG Clark, IB Levitan & MJ Shipston, *J Biol Chem* **276,** 7717–7720 (2001).

[Need the list of references for the "how we know HH were right" discussion.]

:

The problem of setting the numbers of each kind of ion channel emerged in attempts to make quantitative models of individual neurons in the stomatogastric ganglion. For a recent overview of the STG, emphasizing its role as a model system for studying network dynamics, see Marder & Bucher (2007). These models reached a very high degree of sophistication, as described in the series of papers by Golowasch & Marder (1992), Buchholtz et al (1992) and Golowasch et al (1992). The basic idea of regulating the number of ion channels via feedback from the electrical activity of the cell was described by LeMasson et al (1993); see Abbott & LeMasson (1993) for a more complete account. Dramatic experimental evidence for "self–tuning" of channel numbers came (quickly) from Turrigiano et al (1994). For feedback mechanisms with sensitive to multiple time scales, see Liu et al (1998).

Buchholtz et al 1992: Mathematical model of an identified stomatogastric ganglion neuron. F Buchholtz, J Golowasch, IR Epstein & E Marder, *J Neurophysiol* **67,** 332–340 (1992).

Golowasch & Marder 1992: Ionic currents of the lateral pyloric neuron of the stomatogastric ganglion of the crab. *J Neurophysiol* **67,** 318–331 (1992).

Golowasch et al 1992: The contribution of individual ionic currents to the activity of a model stomatogastric ganglion neuron. J Golowasch, F Buchholtz, IR Epstein & E Marder, *J Neurophysiol,* **67,** 341–349 (1992).

LeMasson et al 1993: Activity–dependent regulation of conductances in model neurons. G LeMasson, E Marder, & LF Abbott, *Science* **259,** 1915–1917 (1993).

Marder & Bucher 2007: Understanding circuit dynamics using the stomatogastric nervous system of lobsters and carbs. E Marder & D Bucher, *Annu Rev Physiol* **69,** 291–316 (2007).

Abbott & LeMasson 1993: Analysis of neuron models with dynamically regulated conductances. LF Abbott & G LeMasson, *Neural Comp* **5,** 823–842 (1993).

Liu et al 1998: A model neuron with activity–dependent conductances regulated by multiple calcium sensors. Z Liu, J Golowasch, E Marder & LF Abbott, *J Neurosci* **18,** 2309–2320 (1998).

Turrigiano et al 1994: Activity–dependent changes in the intrinsic properties of cultured neurons. G Turrigiano, LF Abbott & E Marder, *Science* **264,** 974–977 (1994).

Need references to second generation of experiments on mRNA levels. Maybe some pointers to work on networks??

C. The states of cells

Cells have internal states. Sometimes these states are expressed in a very obvious way, even to external observers, as when we see the alternating black and white stripes of a zebra. In other cases, the states are hidden, as when a neuron stops responding to a constant external stimulus, but then rebounds when the stimulus is removed; the amplitude of the rebound reflects the initial amplitude of the stimulus, which must have been stored in some internal state, separate from the output. In these two examples, we also see that these internal states can be discrete or continuous. In many cases, the states of cells are known to be encoded by the concentrations of particular, identifiable molecules, and these concentrations in turn reflect a balance of multiple kinetic processes. If we try to transcribe these qualitative ideas simply into quantitative models, we will find that the states of cells depend on parameters. Most obviously, these states will depend on absolute concentrations, and there is a widespread suspicion that absolute concentrations are highly variable, making them poor candidates for the markers of cellular state. More generally, it would seem that, unless we are careful, states will depend sensitively on parameters, providing another example of the problem of fine tuning vs. robustness that we have been discussing.

In this section we will look at the issue of fine tuning in a variety of biochemical and genetic networks. Historically, these discussions have been independent of the earlier work on protein folding or ion channel dynamics, although I hope to make clear that the conceptual questions are the same. We'll start with the problem of adaptation to constant sensory signals, and move to more complex examples in the cell cycle and embryonic development.

When you tie your shoes in the morning, you can feel the pressure against the skin of your foot, but very quickly this sensation dissipates. When you step outside on a bright summer morning, you are aware of the light, but soon everything looks normal, and you would have trouble reporting accurately the absolute light level. These are examples of sensory or perceptual adaptation, in which we gradually become unaware of constant stimuli, while maintaining sensitivity to small changes in these incoming signals. One of the first things discovered when it became possible to record the signals propagating along individual nerve fibers is that this adaptation occurs, at least in part, in the response of the single cells that first convert sensory inputs into electrical signals, as shown in Fig 106. Further, as we have seen in the discussion of bacterial chemotaxis (Section II.B), adaptation occurs even in the sensory systems of single celled organisms. As we will discuss in connection with the problems of information transmission in neural coding (Section IV.C), adaptation can be a rich and complex phenomenon, being driven not just by constant background signals, but also by the statistical structure of fluctuations around this background.

In the simplest case, where adaptation consists of reducing the response to constant signals while maintaining

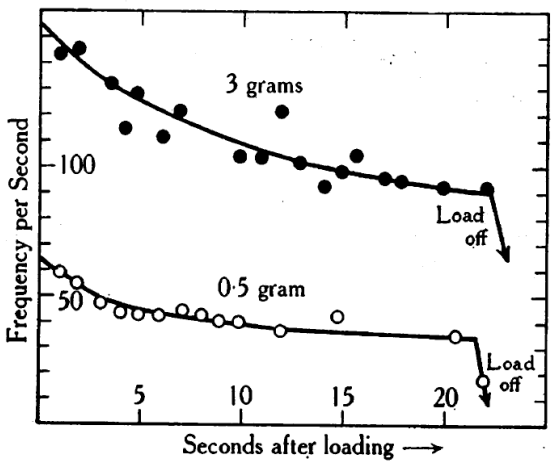

FIG. 106 The original experiments demonstrating adaptation in the response of single sensory neurons (here from the muscle spindle) exposed to constant stimuli (weights), from Adrian & Zotterman (1926a).

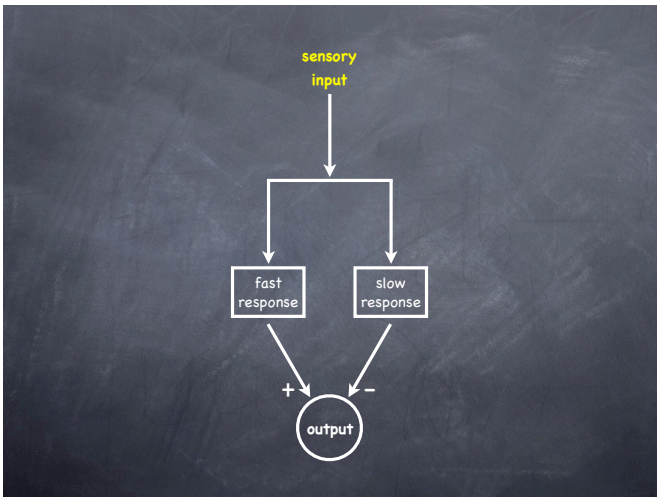

FIG. 107 A schematic of the mechanisms underlying sensory adaptation. The branch which generates fast responses insures that sudden changes in input will be transduced faithfully. The branch with the slower response causes a gradual decay of the output in response to constant inputs. To have truly zero response to constant input requires that the two branches be perfectly balanced.

sensitivity to small transient changes, there is a natural schematic model (Fig 107) in which a rapid positive response to the sensory input is cancelled by a slower negative response. In several systems we can identify the molecular or cellular components that correspond to these different branches, and we will discuss the example of bacterial chemotaxis in detail. For the moment, however, our concern is more general. If adaptation is accomplished through some pathway that is independent of the basic response to incoming stimuli, then the 'gain' of the two pathways are set by independent parameters. If we want the responses to constant inputs to be small, then these two gains must be very similar, so that they nearly cancel. In particular, if we want truly zero response to constants—zero net gain at zero frequency—then the signals passing through the two branches need to cancel exactly, and this seems to require fine tuning of the parameters.

Before saying that we have found a problem, we should examine the precision of cancellation that is actually required. In the example of the fly photoreceptors, discussed in Section I.A, we saw that the system acts as a nearly ideal photon counter up to rates of $\sim 10^5$ photons/s. If the response to a single photon lasts (at its shortest) ~ 10 ms, this means that cell is effectively counting up to ~ 1000. But, as we noted, single photon responses are on the order of a few milliVolts, so if things just add up the voltage across the cell membrane would have to change by several Volts, and this isn't going to happen—something like $90 - 99\%$ of this response needs to be cancelled in order to fit into the available dynamic

range.

In the case of bacterial chemotaxis, we have seen in Section II.B that adaptation is essential for function (see especially Problem **). Because the cell makes decisions based on the time course of concentrations along its trajectory, having a response to constant stimuli would mean that the cell effectively confuses "things are good" for "things are getting better," and this would impede progress up the gradient of desirable chemicals. Direct measurements of the clockwise vs. counterclockwise rotation of the flagellar motor, as in Fig 47, show that the response to a small step in the concentration of attractant molecules decays to zero, so that adaptation is nearly perfect. Another way of seeing this is if one exposes the cells to concentrations that are exponentially increasing in time, the fraction time the motor spends running clockwise become constant, depending on the rate of exponential increase, rather than rising up to saturation; an example is in Fig 108.

Problem 111: Exponential ramps. Give a problem to work out why Fig 108 makes sense!

If we observe freely swimming bacteria, then we can count the rate at which they initiate tumbles, and see that this also adapts to constant stimuli; Fig 109 shows an unnatural but dramatic example, in which a population of bacteria is suddenly exposed to milliMolar concentrations of aspartate, starting from zero background concentration. Tumbling is almost completely suppressed

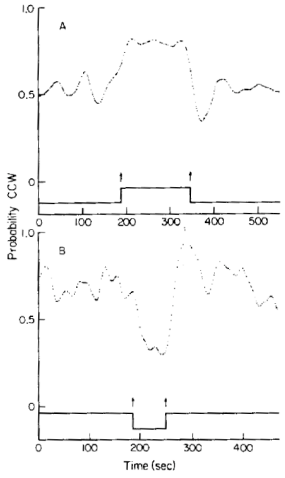

FIG. 108 Response of *E coli* to exponentially increasing (top) or decreasing (bottom) concentrations of an attractant, from Block et al (1983). Probably needs more explanation.

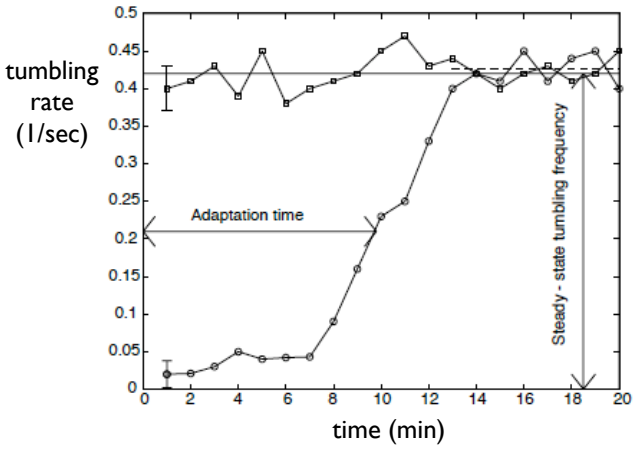

FIG. 109 Experiments on adaptation in a large population of *E. coli* (Alon et al 1999). At time $t = 0$, the population is exposed to a high concentration of an attractive chemical, and as a result the bacteria almost stop tumbling. Over time, they adapt, and the average rate of tumbling approaches the steady–state value observed in the absence of stimuli.

for nearly ten minutes, but eventually recovers to within $\sim 10\%$ of its initial rate, despite the fact that the initially saturating stimulus continues to be present [show an earlier figure of this flavor from Berg or Koshland?].

To understand how it's possible to achieve near perfect adaptation without fine tuning of parameters (as one might have thought from Fig 107), we have to dig into the details of the molecular mechanisms involved. In Section II.B we outlined the fast events involved in the "positive" part of the chemotactic response (Fig 48). To review briefly, receptor molecules on the cell surface form a complex with the enzyme CheA (a kinase), held together by a scaffolding molecule CheW. The complex is in equilibrium between the active (CheA*) and inactive (CheA) states, and this equilibrium is shifted by binding of attractant or repellent molecules to their receptors; for attractants, binding shifts the equilibrium toward the inactive state. The active kinase CheA* phosphorylates the protein CheY, which can diffuse through the cell from the receptor complex to the flagellar motor, where it binds and favors clockwise rotation, driving the tumbling motion of the cell; the action of the kinase is opposed by a phosphatase, CheZ. Thus, an increase in the attractant concentration drives the kinase toward its inactive state, reducing the rate of phosphorylation of CheY; the continued action of the phosphatase results in a reduction of the CheY–P concentration, and this reduces the probability of tumbling. This whole pathway is extraordinarily sensitive, responding reliably to individual molecules as they bind to their receptors.

How does the extremely sensitive response of the chemotactic system get cancelled when stimuli are maintained at constant levels? In addition to binding the

chemoattractant or repellent molecules, the receptors can be modified by covalent attachment of methyl groups. Much as with ligand binding, these modifications shift the equilibrium between active and inactive conformations of the kinase CheA—binding of attractants favors the inactive state, addition of methyl groups favors the active state. The key point is that the active kinase not only phosphorylates CheY, leading to clockwise rotation of the motor, it also phosphorylates CheB, and then CheB–P removes methyl groups from the receptor. Thus, when an attractant lowers the activity of the kinase, it also allows more methyl groups to be attached, driving the activity back toward its original level—adapting.

Although the methylation system provides a pathway to cancel the effect of the immediate response to sensory inputs, it isn't clear that this cancellation should be anywhere near exact. In general, one would need to tune the activity of the methylation and de–methylation enzymes to make sure that their effects exactly balance the direct response to sensory input. So, this system provides an example of our general problem of fine tuning, as emphasized by Barkai and Leibler. In addition to identifying the problem, they proposed that one can evade this need for fine tuning by assuming that the de–methylation enzyme CheB only recognizes the active state of the receptor–kinase complex, and ignores the inactive conformation. If this is true, one doesn't even need the phosphorylation of CheB in order to close the feedback loop.

To see how the Barkai–Leibler scheme works, let's imagine that the whole receptor complex, which might include a cluster of several receptor molecules, switches as a whole between active an inactive states. There is some free energy difference ΔF between these states, and there are two contributions to this difference—one from the binding of attractants, and one from methylation. Assume that the contribution of the methyl groups is additive, and that the contribution from ligand binding has some arbitrary dependence on ligand concentration c (which we could work out from a model like that in Fig 110; see Problem 112 [check] below). Then the number of active enzymes is given by

$$A^* = \frac{A_{\rm total}}{1+\exp\left[F_L(c) - n_{\rm M}\Delta_{\rm M}\right]}, \tag{552}$$

where n_M is the number of methyl groups per receptor complex. This number reflects a balance between the activities of CheR and CheB, so we can write schematically

$$\frac{dn_M}{dt} = V_R - V_B, \tag{553}$$

where V_R and V_B are the 'velocities' of the methylation and de–methylation enzymes, respectively.

The key assumptions suggested by Barkai and Leibler are that CheR is running at some maximal rate, limited by its internal dynamics and not by the availability of substrate, while the velocity of CheB does depend on the availability of its substrate A^* according to some function $f(A^*)$ that we don't need to specify. Then

$$\frac{dn_M}{dt} = V_R^{\rm max} - V_B^{\rm max} f(A^*). \tag{554}$$

In order to reach steady state ($dn_M/dt = 0$), we must have

$$A^* = A_0^* = f^{-1}(V_R^{\rm max}/V_B^{\rm max}), \tag{555}$$

independent of the ligand concentration c. Thus all steady states in the system must have the same level of activation of the kinase, hence the same level of phosphorylation of CheY and the same rate of tumbling. These steady states at varying c are not identical—they involve different levels of methylation—but they have the same functional output.

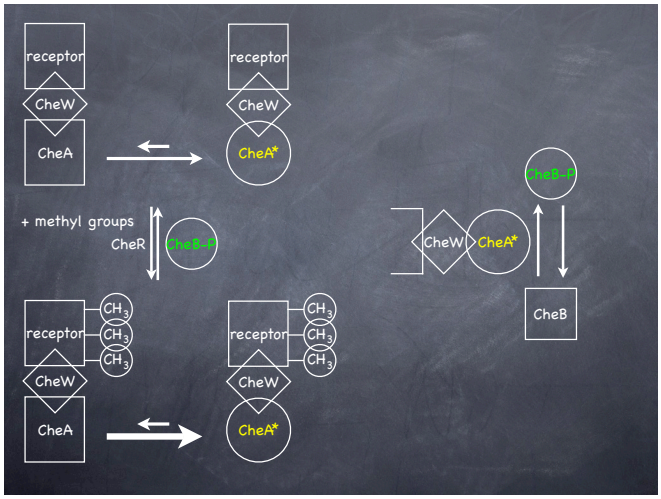

FIG. 110 Methylation of the receptors allows for adaptation of the chemotactic response. At left, addition of methyl groups acts, similarly to ligand binding, as an allosteric effector, shifting the equilibrium between the active and inactive states of the kinase CheA; the schematic is meant to indicate that there are multiple methylation sites. At right, the feedback loop is closed by having the active kinase CheA* trigger activation of the de–methylation enzyme CheB. Need to redraw to remind that methylation is working opposite to the effects of an attractant binding.

Problem 112: Allosteric model for chemotactic receptors. [check for earlier problem about this ...] The schematic in Fig 48 is equivalent to a Monod–Wyman–Changeaux model (all relevant pointers) in which the whole complex of the receptor, CheW and CheA has two states, and the equilibrium is shifted by binding of the attractant molecule. In Fig 110, attachment of methyl groups also shifts this equilibrium, but the binding and unbinding of these groups is part of an energy–yielding reaction, and so doesn't have to obey detailed balance. Show that, nonetheless, these schematics generate Eq (552), which has a decidedly Boltzmann form. Why

does this work? What would change if groups or clusters of N receptor complexes were tied together, and forced to all be in the same activation state?

If the scenario sketched here is correct, then we should be able to test it by manipulating the activity of the methylation and de–methylation enzymes, using the modern tricks of molecular biology to modify the genome of *E coli*. To begin, one can replace CheB with a mutant form which cannot be phosphorylated; adaptation still works, and still is nearly perfect, suggesting that phosphorylation is not the key step in closing the feedback loop. Then one can delete the normal CheR gene and replace it with a plasmid which carries the CheR coding region under the control of a promoter that responds to external signals. In this way one can generate roughly 100–fold variations in CheR expression levels, from half the normal level to 50× over–expression, as in Fig 111. Throughout this range, adaptation to large inputs (as in Fig 109) is within $\sim 10\%$ of being exact. Although the mean rate of tumbling to which the system adapts, as well as the time scale of this adaptation, depends on the amount of CheR in the cell, the fact that this rate is independent of input concentration does not. Are there experiments that look at adaptation in response to smaller signals? Maybe from Sourjik?

There is a lot of evidence that the methylation level of the receptors really is the molecular representation of the cell's adaptation state. As such, we might have expected that over– or under–expressing the enzyme that carries out the methylation reaction would shift the actual state of the system, and this would show up as a change in the output. In the model considered here, however, this last expectation is violated. The absolute level of kinase activity, and hence the absolute tumbling rate, does indeed change when we change the expression of CheR. But Fig 111 shows us that the average steady state response to an applied step in attractant concentration remains zero, independent of the CheR level. Thus, the precision of the balance between the processes responsible for excitation and adaptation does not depend on fine tuning of the underlying kinetics.

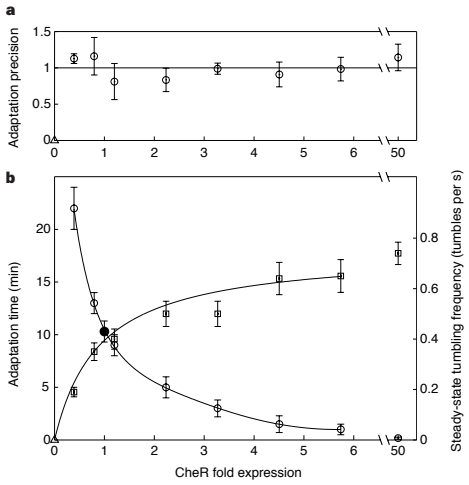

FIG. 111 Chemotactic responses in the presence of varying amounts of CheR, from Alon et al (1999). At the top, 'adaptation precision' is measured as the ratio of the mean tumbling rates in the presence and absence of 1 mM aspartate (as in the experiments of Fig 109). The actual tumbling rates and the time required to reach steady state after sudden exposure to 1 mM aspartate are shown in the bottom panel.

Problem 113: Calcium driven adaptation in neurons. Consider a neuron that generates spikes at rate r. Let's assume the response to external inputs I involves this rate relaxing toward some steady state,

$$\tau \frac{dr}{dt} = r_{\rm max} f(I, [{\rm Ca}]) - r, \tag{556}$$

where we note explicitly that the rate depends both on the inputs and on the intracellular calcium concentration. Write an equation for the dynamics of [Ca], assuming that each spike brings in a fixed number of calcium ions, and that there is a pump which extrudes the ions at some opposing rate. The pumping rate must depend on the concentration, but for the moment take this dependence as some unknown function $V_{\rm pump}([{\rm Ca}])$. Find equations that describe the steady state of this system. Are there conditions on $V_{\rm pump}([{\rm Ca}])$ that lead to a steady state spike rate that is independent of the input I? If the input changes suddenly, does the spike rate still respond? Explain how this relates to the discussion of chemotaxis given here.

Are we done? I think there is still more to this problem. To begin, the fact that motor output is an extremely steep function of the CheY–P concentration (pointer) means that successful adaptation requires more than just a constant level of CheY–P in steady state, independent of the input signal—this level actually has to fall into a very narrow range, or else the cells will be always running or tumbling. The parameters which determine the steady state level of CheY–P are independent of the properties of the motor, which determine the functional operating range for this concentration. This seems like the same sort of balancing problem that Barkai and Leibler were worried about, but in a different part of the system, where their solution has no obvious analog. [Has somebody worried about this?]

Next, you should also be a little suspicious about the simple equations above. At best, they are some sort of mean field theory in a system where fluctuations could be important. Also, while it's plausible that CheB recognizes CheA* as opposed to CheA, one might worry that the rate of removing methyl groups depends on how many are there (especially if that number goes to zero!). There

must be some regime in which the simple argument is right, but we need a more honest calculation. [several groups have tried this; look closely at Wingreen et al, and check for others, to decide what to say here.]

Finally, although one can manipulate the *E coli* genome to change expression level of individual proteins by large factors, the many protein components of the chemotactic system are encoded on just two operons, which means that the expression of the different components is tightly coupled under normal conditions [be sure to have talked about operons before this, or maybe this is really a good place to introduce the idea?].

The *mocha* operon encodes CheA and CheW, along with the flagellar motor proteins, and the *meche* operon encodes CheR, CheB, CheY and CheZ, along with two classes of receptor proteins. Recent experiments indicate that there is covariation even between the expression levels of CheA and CheY, suggesting that the cell can in fact control at least the relative concentrations of these proteins fairly precisely. Further, there is direct evidence that tight correlation between protein concentrations actually improves chemotactic performance, as shown in Fig 112.

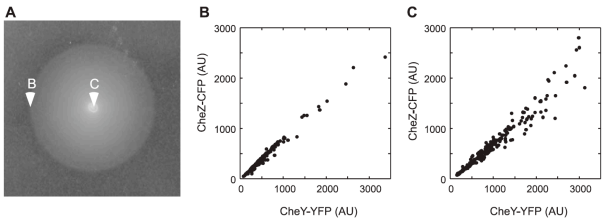

FIG. 112 Better chemotactic performance is associated with correlated fluctuations in protein levels, from Løvdok et al (2009). At right, *E coli* swarm outward toward attractants. Cells have been engineered to express CheY and CheZ only under the control of a promoter induced by external signals. If we select cells from regions B or C of the swarm, we see that the cells which have been efficient (B) have tightly correlated variations in the two protein levels, while cells that have been inefficient (C) have weaker correlations. Thus, selection for chemotactic efficiency will drive down the *relative* fluctuations in expression levels, even there is substantial tolerance for variation in the absolute levels.

Problem 114: Balancing CheY and CheZ. Take the students through a model in which it becomes clear why variations in the relative levels of CheY and CheZ are detrimental for chemotaxis, thus making sense out of Fig 112.

It is interesting to compare the problem of robustness vs. fine tuning in the case of chemotaxis with what we learned in the case of ion channels (Section III.B). For ion channels, function really does depend sensitively on the number of copies of the different proteins in the network, and neurons have evolved control mechanisms that use their functional output (or a near surrogate) to control these copy numbers. Importantly, there are many ways to achieve the same function, so it is not the number of copies of each component that is tuned, but rather some possibly complex combinations of these quantities. For chemotaxis, the message of the experiments in Fig 111 is that large variations in the copy number of just one component can be tolerated, pointing toward networks that are intrinsically insensitive to this parameter variation rather than any hidden control or tuning mechanisms. This suggests that one system is tuned, and the other is robust.

On the other hand, Fig 112 shows that, as with ion channels, the relative copy numbers of the proteins in the chemotaxis network *are* controlled, and that this control contributes to function. Experiments more directly analogous to Fig 111 have now been done in stomatogastric ganglion neurons, and one finds that there are control mechanisms which can compensate for over–expression of particular channel types by changing the expression levels of other channels see Fig 113. Perhaps surprisingly, these compensation mechanisms are triggered even if the first channel is non–functional and hence doesn't effect the electrical output, suggesting that the there are signals internal to the transcriptional and translational networks which encode something about the correct, functional operating point of the system. This could be a much more general phenomenon.

Before moving on, there is also a somewhat philosophical point to be made about the mechanism of robustness in chemotaxis, or perhaps even about the idea of robustness itself. If we expect the function of a biochemical network to be robust against parameter variation, this robustness must be a property of the network topology—which nodes (molecules) are connected by arrows (reactions). In the specific model considered by Barkai and Leibler, for example, it is essential that CheB acts on CheA* as a substrate, but not on the inactive CheA—what is important in this case is the *absence* of an link in the network connecting CheB with CheA.

The particular links that appear (or don't appear) in a biochemical network reflect the specificity of the various enzymatic and protein–protein binding reactions. Substrate specificity is a classical topic in biochemistry, and much of what we understand about this topic was learned through painstaking experiments on purified samples of particular enzymes. The ideas of robustness emerged at a time when the community started to wonder if there wasn't something a bit hopeless about the overall project

of this classical biochemical approach. While one can study individual enzymes in detail, many interesting biological functions emerge from networks with many interacting proteins engaged in dozens if not hundreds of individual reactions. Further, the conditions inside the living cell may be far from those that we can reproduce in a test tube. How then could we ever study every one of the relevant reactions, under the right conditions? Seen in this light, it just doesn't seem plausible that the accumulated biochemical knowledge will "add up" to give us an understanding of how cells function as complete systems. Robustness was one of several ideas offered as an alternative—if Nature has selected networks that are robust to parameter variations, then the (already somewhat hopeless) project of measuring all these parameters could safely be abandoned. But because network topology is an encoding of substrate specificity, we can't really brush aside all of classical biochemistry. Indeed, the example brought forward by Barkai and Leibler is one in which the biochemistry is subtle, with one protein recognizing different conformations of another. At the end of the day, then, biochemistry has its revenge—robustness may be an emergent, system level property, driven by network topology, but this topology is an expression of the underlying, detailed biochemistry.

[I'd like put here a discussion of the work by Tang and coworkers on the cell cycle. The idea is that not just states, but trajectories through the space of states, are robust. Need to sort through the papers for details. Should also discuss the results from the Siggia/Cross collaboration on the cell cycle, in particular what makes the decision to go from one state to the next reliable and irreversible.]

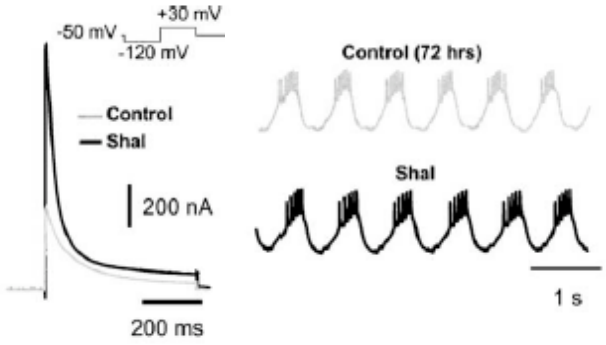

FIG. 113 Responses to over–expression of a channel, from Maclean et al (2003). At left, injecting mRNA for the A current channel (see Table I) produces, after 72 hrs, an increase in the current that flows when the voltage is stepped through the range expected to activate this channel. This shows that injecting the mRNA really does result in more channels being synthesized and inserted into the membrane. At right, a demonstration that this increased number of channels (in the "Shal" trace) does not perturb the basic pattern of activity (seen in the control). This is possible only because the cell compensates by increasing the expression levels of other channels.

Another example of "robust" output from a biological network is that almost everyone you know was born with five fingers on each hand. In insects, one can count even more instances in which discrete pieces of the body are arranged in a repetitive pattern, from the segments of the body itself (as in the beautiful caterpillar shown in Fig 114) to the hairs or bristles on the body surface; essentially every member of a particular species has the same number of body segments, the same number of hairs, and even the positions of the hairs are identifiable from individual to individual. It is not at all obvious how this level of reproducibility is achieved.

FIG. 114 Insects provide many examples of repeated, reproducible structures visible on the outside of the body. Image of tiger moth caterpillar from http://www.hsu.edu/content.aspx?id=7435.[probably should take our own picture; also, maybe another panel about segments, or bristles?]

Broadly speaking we can distinguish two classes of explanations for the reproducibility of pattern formation ('morphogenesis') in the embryo. In the first kind of explanation, the organism works to set the initial conditions and boundary conditions very precisely, and each step in the process has been tuned to minimize noise. Patterns then develop in a reproducible fashion in the same way that accurate clocks continue the same time even though tick independently. Alternatively, it is possible that noise and errors abound, but that there are error correction mechanisms that pull the pattern back to its ideal structure. Of course, it is also possible that both scenarios are correct: nature has selected for systems with minimal noise, and taken care to control the conditions of development, but error correction mechanisms still are needed to deal with the vagaries of a fluctuating environment.

To appreciate why the observations of reproducibility in morphogenesis are so puzzling, we need to review some of the basic mechanisms by which patterns form in the developing embryo. We will also need to check our qualitative impressions of reproducibility against quantitative data. Let's start with the background.

We recall that embryos start as just one cell, the fertilized egg, and then there are multiple cell divisions. Every one of these cells (as in our adult bodies) has the same DNA, assuming that nothing has gone wrong. What makes the different cells different is that they "express" different genes. The genes code for proteins, but not all of the proteins are made in all cells; the reading of the code to make the proteins is called the expression of the genes, as we have discussed before. Importantly, the regulation of gene expression is not just the flipping of a switch sometime in development, but rather something that all cells (from neurons in our brain down to bacteria) are doing all the time. Embryos come in all shapes and sizes throughout the animal kingdom, but for various reasons people have focused on a few model systems, and we will do the same.

The fly embryo is an interesting model system for many reasons. One is that there is a well developed genetics for fruit flies (the species *Drosophila melanogaster*), made possible not least by their rapid growth and reproduction. Embryonic development itself is rapid as well, leading from a fertilized egg to the hatching of a fully functional maggot (the larvae of flies, like caterpillars for butterflies) within 24 hours. All of this happens inside an egg shell, so there is no growth—pattern formation occurs at constant volume. The egg is $\sim 1/2\,$mm long, so one starts with one rather large cell, which has one nucleus. In the maggot there are $\sim 50,000$ cells. For the first three hours of development, during which the "blueprint" for the body plan is laid out, something special happens: the nuclei multiply without building walls to make separated cells. Thus, for about three hours, the fly embryo is close to the physicists idealization of a box in which chemical reactions are occurring, with the different molecules free to move from one end of the box to the other (perhaps even by diffusion, although this is a more subtle question).

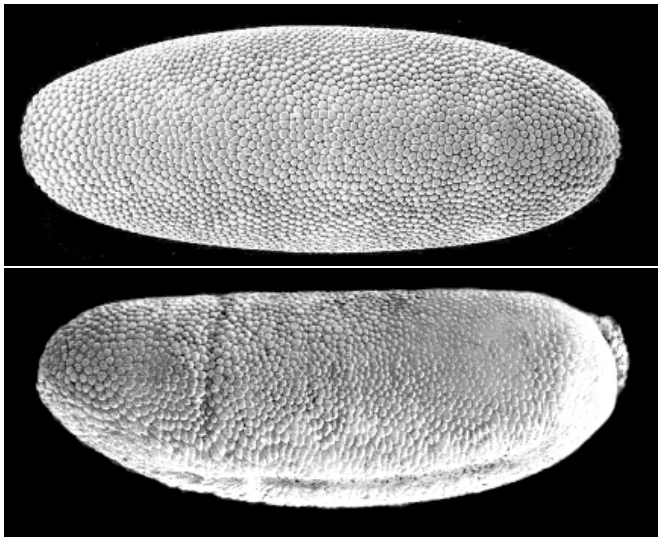

FIG. 115 Electron micrographs of a *Drosophila* embryo in cycle 14, before (top) and after (bottom) gastrulation. Note, in particular, the cephalic furrow roughly one third of the distance from the left in the bottom image. Micrographs taken by EF Wieschaus.

The duplication of the nuclei is more or less synchronous for the first 13 mitotic divisions, or nuclear cycles, which is visually quite striking. During cycles 8 through 10, almost all of the nuclei move to the surface of the egg, where they form a fairly regular two dimensional lattice; conveniently, with all the nuclei at the surface of the egg, we have a much better chance to "see" what is going on (see Figs **??** et seq). With each subsequent cycle, this lattice dissolves and reforms. With cycle 14, the synchronous duplication of nuclei stops, and there is a pause while the embryo builds walls between the nuclei to make separate cells. If you stop the action at this point and take an electron micrograph of the embryo, what you see is at the top in Fig 115. If you count, you'll find that there are ~ 6000 cells on the surface. This is smaller than 2^{13}, but thats because not all of the nuclei make it to the surface; some stay in the interior of the embryo, probably not by accident since these become cells with special functions. Notice that all the cells look pretty much alike. If instead of stopping at this point, we wait just 15 minutes more, we see something very different, shown at the bottom in Fig 115. Notice that there is a vertical cleft, about one-third of the way from the left edge of the embryo. This is the "cephalic furrow," and defines which part of the body will become the head. There is also a furrow along the bottom of the embryo, which is where the one layer of cells on the surface starts to fold in on itself so that you can have two "outside" surfaces (think about the inside and outside of your cheek, both of which are outside of the body from the topological point of view we are not simply connected!), a process called "gastrulation."

Its not just that the embryo breaks into a head and a non–head. In fact there are many different pieces to the body, usually called segments, as noted above. The obvious question is how the cells at different points in the embryo know to become parts of different segments. The answer is quite striking, and one of the great triumphs of modern biology. Long before cells start moving around and making the three dimensional structures that one sees in the fully developed organism, there is a "blueprint" that can be made visible by asking about the expression levels of particular genes. A now classic set of genetic experiments showed that the number of genes that are relevant in these early patterning events is small, on the order of 100 out of the roughly 25,000 genes in the whole fly genome; if we focus on the pattern along the anterior–posterior axis of the embryo, the number of relevant genes is less than 20. Most of these genes code for transcription factors that control the expression of other genes.

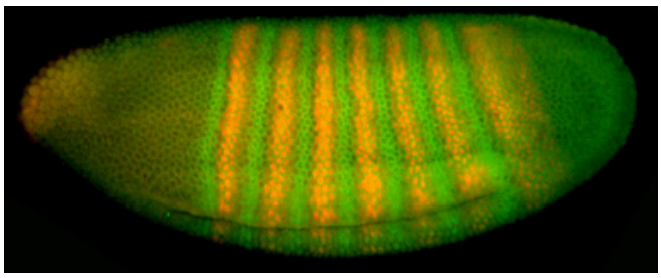

FIG. 116 [...] Thanks again to EF Wieschaus for these images.

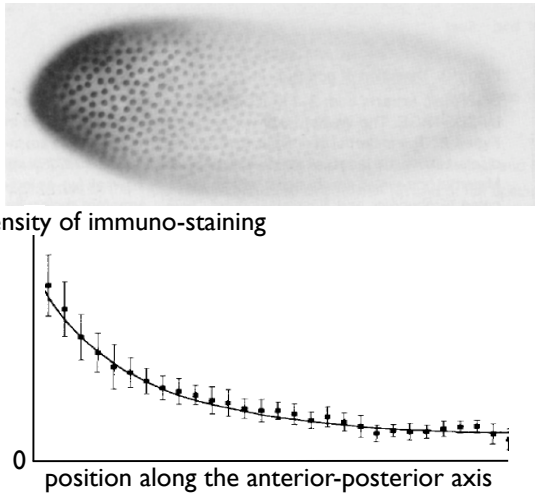

FIG. 118 Antibody staining for the protein Bicoid in the early *Drosophila* embryo, from the original experiments by Driever & Nüsslein–Vollhard (1988a). The plot at the bottom represents means and 2× standard deviations from ten embryos; units of staining intensity are arbitrary. [Are these errors bars realy 2× the standard deviation, or just ± the standard deviation? Might have to ask the authors—for a preliminary result, one can't complain about a factor of two, but I want to get it right!]

Suppose we stop that action in the embryo at cycle 14 and measure the concentration of two of these key proteins. One way to do this is to make antibodies against the protein we are interested in, and then make antibodies against the antibodies, but before using the secondary antibodies we attach to them a fluorescent dye molecule. So if we expose the embryo first to one antibody (which should stick to the protein we are interested in, and not anywhere else, if were lucky) and to the other, we should have the effect of attaching fluorescent dyes to the protein we are looking for, and hence if we look under a microscope the brightness of the fluorescence should indicate the concentration of the protein (not obvious if this relationship is quantitative; hold that question). One such experiment is shown in Fig 116. Evidently the concentration of the proteins varies with position, and this variation corresponds to a striped pattern. The stripes should remind you of the segments in the fully develop animal, and this is actually quite precise. Mutations that move the stripes around, or delete particular stripes, have the expected correlates in the pattern of segmentation. To illustrate this point, we can blow up corresponding pieces of this image and the electron micrograph above, showing the cephalic furrow (Fig 117); hopefully you can see how the furrow occurs at a place defined by the locations of the green and orange stripes. At the moment the names of these molecules don't really matter. What is important is to realize that the macroscopic structure of the fully developed organism largely follows a blueprint laid out within about three hours after fertilization, and that this blueprint is "written" as variations in the expression level of different genes. Furthermore, we know which genes are the important ones, and there aren't too many of them.

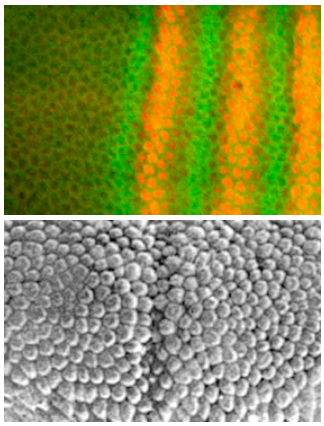

FIG. 117 A combination of Figs 115 and 116, emphasizing that the cephalic furrow occurs along a single line of cells that can be identified from the pattern of pair rule gene expression.

We have pushed the problem of pattern formation in the embryo back to spatial variations in the pattern of gene expression, but how do these arise? You could imagine, as Turing did, that these patterns reflect a spontaneous breaking of symmetries in the egg. This, for better or worse, is not how it works. When the mother makes the egg, she places the mRNA for a handful of proteins at cardinal points. For example, there is a protein called Bicoid for which the mRNA is placed at the end that will become the head; importantly, the mRNA is attached to the end of the egg, not free to move. Once the egg is laid, translation of this mRNA begins, and the resulting Bicoid protein is free to move through the embryo. If we

use the same trick as above and stop the action, labeling the embryo antibodies against the protein, we see images like those in Fig 118. Evidently there is a rather smooth gradient in the concentration of Bicoid protein, high at one end and low at the other. A cell sitting at some point in the embryo thus can "know" where it is along this long (anterior–posterior) axis by measuring the Bicoid concentration. This is an example of the very general idea of "positional information" in the embryo.

Since Bicoid is a transcription factor, it provides an input signal to a whole network of interacting genes, and this network can (if we speak colloquially) interpret the positional information, ultimately driving the emergence of the beautiful striped patterns as in Fig 117. We'll look in more detail at how this happens, but for now let's try to sharpen our questions about reproducibility.

Measurements on the profile of Bcd concentration show rather decent agreement with an exponential decay, as was noted already in the very first experiments (Fig 118), so that

$$c(x) \approx c_0 e^{-x/\lambda}, \tag{557}$$

where x is measured from the anterior end of the egg. Suppose, then, that the cephalic furrow is placed at the point where the Bcd concentration reaches some threshold value θ_{cf}. The position of the cephalic furrow is then

$$x_{cf} = \lambda \ln(c_0/\theta_{cf}). \tag{558}$$

Thus, if c_0 changes by $\sim 10\%$, the location of the furrow would shift by $\delta x_{cf} \sim 0.1\lambda$. Experimentally, modern experiments show that $\lambda \sim 100\,\mu$m, and the location of the cephalic furrow is reproducible with a standard deviation of $\sim 1\%$ of the length of the embryo, or $\sim 5\,\mu$m in absolute length. In fact, one can look at other positional markers, such as the locations of peaks or troughs in the striped patterns of expression for the "pair rule" genes in Fig 116, and these are all reproducible at the $\sim 1\%$ level, as shown in Fig 119. Thus, taken at face value, if the Bcd profile provides the basic "map" of position along the anterior–posterior axis, then the absolute concentration of Bcd, c_0, would have to be reproducible to better than $\sim 10\%$ from embryo to embryo in order to generate the observed reproducibility of these patterns. This problem exists even before we ask how to maintain constant proportions in the face of variations in the overall size of the embryo.

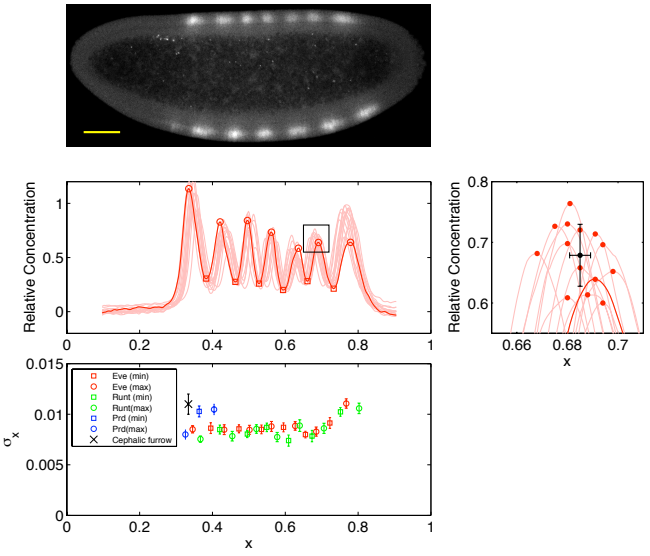

FIG. 119 Reproducibility of various spatial markers along the anterior–posterior axis in the early *Drosophila* embryo, from Dubuis et al (2011). At top, fluorescent antibody staining of the protein Eve; scale bar is $50\,\mu$m. Middle, normalized spatial profiles of the fluorescent intensity in 14 embryos; the darker red line is the embryo shown at the top. At right, a small region is blown up to show the variability of the peak; error bars show standard deviations of position and amplitude. Bottom, standard deviations of position for peaks and troughs of several gene expression profiles, as well as for the position of the cephalic furrow measured in live embryos.

I think that, when people started to think about this problem quantitatively, it seemed implausible that the reproducibility of embryonic development would depend on controlling absolute concentrations with 10% accuracy. One the other hand, the paper which first characterized the spatial profile of Bicoid protein actually reported data on the variations across embryos (Fig 118), and the results are roughly consistent with $\sim 10\%$ reproducibility, at least near the anterior end of the embryo.

Let's take seriously the simplest possible model for the spatial profile of Bicoid (Bcd), described above in words: the mRNA placed by the mother acts (as it is translated) as a source, the Bcd protein diffuses through the embryo, and the protein is also degraded by some first order reaction. If we simplify and think of the system as just being one dimensional (along the anterior–posterior axis), then the concentration $c(x,t)$ should obey

$$\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2} - \frac{1}{\tau}c(x,t), \tag{559}$$

where τ is the lifetime of the protein against degradation. The boundary conditions are

$$-D\frac{\partial c(x,t)}{\partial x}\bigg|_{x=0} = R, \tag{560}$$

$$\frac{\partial c(x,t)}{\partial x}\bigg|_{x=L} = 0, \tag{561}$$

where R is the strength of the source at $x = 0$ and the last condition states that there is no flux out of the other end of the embryo. If we imagine that development is slow enough for the system to come to steady state, and that the embryo is long, the concentration profile becomes

$$c_s(x) = \frac{R\tau}{\lambda}e^{-x/\lambda}, \qquad \lambda = \sqrt{D\tau}. \tag{562}$$

Problem 115: Details of the Bicoid profile.

(a.) What are the units of concentration in one dimension? Show that, with this proper choice of units, R is the number of Bcd molecules being translated per second.

(b.) Derive the steady state solution in Eq (562). What is the precise criterion for the embryo to be long enough that this is approximation is accurate?

(c.) At this writing, there is controversy about whether the Bcd profile really reaches steady state during the early stages of development. Although this is an experimental question, we can ask what the simplest model predicts. Intuitively, there is some time scale t_* do you expect the solution of Eq(559) reaches steady state; how does this time scale relate to the other parameters of the problem? Answer this without doing any detailed calculations, and think about how your intuition might go astray.

(d.) Try to do a more detailed calculation to address the approach to steady state. It is useful to assume from the beginning that L is large [in the sense of part (b.)], and to replace the boundary condition at $x = 0$ with a source in the symmetrized version of the problem,

$$\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2} - \frac{1}{\tau}c(x,t) + 2R\delta(x), \tag{563}$$

where now $-\infty < x < \infty$; be sure you understand why we need a factor of 2 in front of the source term. At $t = 0$, before any protein has been translated, we must have $c = 0$ everywhere. By Fourier transforming in space, show that the exact time dependent solution is

$$c(x,t) = 2R\int_{-\infty}^{\infty}\frac{dk}{2\pi}\frac{e^{ikx}}{Dk^2 + 1/\tau}\left[1 - e^{-(Dk^2+1/\tau)t}\right]. \tag{564}$$

Verify that this approaches $c_s(x)$ from Eq (562) as $t \to \infty$.

(e.) Find a simple closed form for the time derivative of concentration at a point, $\partial_t c(x,t)$. Show that, expressed as a fraction of the local steady state concentration, this derivative peaks at a point $x_* = 2\lambda t/\tau$, and that at this peak $[\partial_t c(x_*,t)]/c_s(x_*) = 1/\sqrt{\pi\tau t}$.

(f.) Suppose we could establish experimentally that, for example, after $t = 1$ hr, at each point x that we can see, $c(x,t)$ changes by less than 1%/min (or $\sim 10\%$ across the time required for the cell cycle). What can you conclude about the parameters of the system, taking the simple model seriously?

We see that this simplest model recovers Eq (557), which was suggested by the data. It gives us an explicit formula for the length constant λ, and tells us (not surprisingly) that the absolute concentration scale c_0 is proportional to the strength of the source—that is, to the rate at which proteins can be translated from the mRNA bound to the anterior end of the embryo. In this simple model, then, if we want c_0 to be reproducible with 10% accuracy, the source strength must also be reproducible. Is it plausible that the mother can count out mRNA molecules, with 10% accuracy, and create an environment in the embryo where the efficiency of translation is similarly well controlled? Alternatively, can we escape from these requirement of fine tuning by moving away from the simplest model?

Suppose that the processes which degrade the Bicoid molecule act not on individual molecules, but on dimers, and these dimers are rare. We then expect that the concentration of dimers will be proportional to the square of the Bcd concentration, and the dynamics become [instead of Eq (559)]

$$\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2} - \frac{1}{\tau c_2}c^2(x,t), \tag{565}$$

where c_2 is the concentration scale for dimer formation. Now the steady state solution must obey

$$\frac{d^2 c_s(x)}{dx^2} = \frac{1}{D\tau c_2}c_s^2(x). \tag{566}$$

Notice that if we look for a solution of the form $c_s(x) = Ax^n$, we have

$$\frac{d^2(Ax^n)}{dx^2} = \frac{1}{D\tau c_2}(Ax^n)^2 \tag{567}$$

$$An(n-1)x^{n-2} = \frac{A^2}{D\tau c_2}x^{2n}, \tag{568}$$

which is solved by $n = -2$ and $A = 6D\tau c_2$. Thus, far from the source, the concentration profile is $c_s(x) = 6D\tau c_2/x^2$ *independent of the strength of the source.* More precisely, to match the boundary condition describing the source at $x = 0$, we have to have

$$c_s(x) = \frac{6D\tau c_2}{(x + x_0)^2}, \qquad x_0 = (12D^2\tau c_2/R)^{1/3}. \tag{569}$$

The strength of the source appears only in x_0; for $x \gg x_0$ this term is negligible, and for large R this condition itself sets in at very small x. In this model, then, just making the source very strong—but not setting the strength precisely—is sufficient to insure that almost the entire concentration profile will be independent of variations in this source strength.

Problem 116: Fill in the arguments leading to Eq (569).

It is interesting that a relatively small change in molecular mechanism makes such a dramatic change in the robustness of the system to variations in parameters. One might object, of course, that here is no free lunch here. While Eq (569) predicts that the Bcd profile is independent of the source strength, the concentration scale is now set by c_2, which has something to do with the dimerization of the molecules. The source strength R depends on how many copies of mRNA the mother places in the egg, but the scale c_2 is determined by more global physical–chemical parameters of the cytoplasm, and perhaps these are easier to control. On the other hand, if degradation is

active, via enzymatic reactions, then τ itself will depend on the number of copies of the enzyme that are present in the embryo. Still, it is interesting to ask whether Nature makes use of such a scheme to reduce the sensitivity of morphogen profiles to variations in the strength of the source. [Should also discuss Bollenbach et al (2005).]

Another approach is to give up on making a single morphogen signal reproducible, and to assume that the embryo makes use of multiple signals, hoping that the dominant sources of variation are in a "common mode" that can be rejected by the network that processes these signals. Several models of this flavor have been suggested [refs: Houchmandzadeh et al (2005), McHale et al (2006), others?]. Need to sort out how much of this is about scaling, and how much about reproducibility.

Problem 117: [Should be able to get one or two problems from the model in the last two paragraphs!]

With all this theoretical background, what can we say about the experimental situation? As noted at the outset, there are hints from the earliest literature that Bicoid profiles in *Drosophila* might indeed be reproducible. We also know that the notion of robustness should not be exaggerated. The success of classical genetics in identifying the components of these networks immediately tells us that the system is *not* resistant to the elimination of single components. More subtly, one of the key experiments in establishing that Bicoid is a primary source of positional information was to change the number of copies of the *bcd* gene; with more (or fewer) copies of the gene in its genome, the mother makes more (or less) mRNA and hence drives the strength of the Bcd source up (or down). In response to these changes, the patterns in the early embryo shift, with the cephalic furrow in particular moving—with higher concentrations of Bicoid, the embryo tries to make a larger head, as shown in Fig 120.[71] These results suggest that the embryo does not engage mechanisms which buffer the Bcd profile against variations in the strength of the source. For the morphogens whose concentration profile varies along the other axis of the embryo, however, there are signatures of the nonlinear degradation mechanism which, as we have seen, can

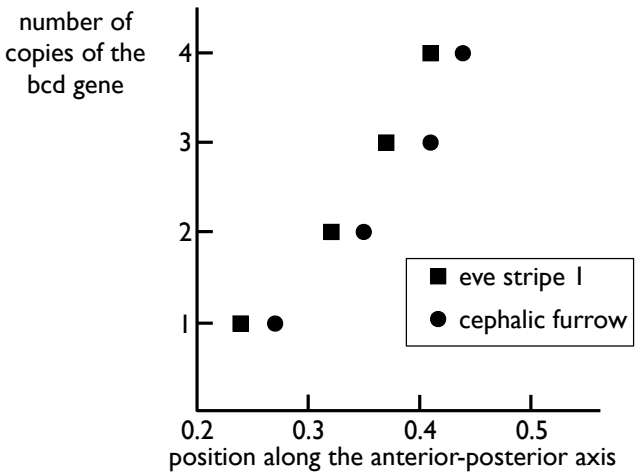

FIG. 120 Variations in the position of spatial markers along the anterior–posterior axis in response to changes in the number of copies of the *bcd* gene, from Driever & Nüsslein–Vollhard (1988b). [explain!]

generate substantial robustness. [Should we have a figure from Eldar et al?]

If there is no buffering, then it really does seem that reproducible outputs require reproducible inputs. Can we see this directly? As discussed in Section II.B, one can genetically engineer flies to express a fusion of Bcd with the green fluorescent protein (GFP), and show that this fusion protein quantitatively replaces the function of the native molecule. Figure 121 shows measurements of the concentration of Bicoid in nuclei from 15 different embryos, using this Bcd–GFP fusion. The raw fluorescence intensity (or the inferred concentration) is plotted vs. position along the anterior–posterior axis for each nucleus. Evidently the variability from embryo to embryo is small, with a standard deviation of less than 20%, and some of this variability can be traced to measurement errors, suggesting that the true variability is $\sim 10\%$ or even less. If the mother has only one copy of the Bcd–GFP gene instead of the usual two, the fluorescence really is cut in half, so again there is no evidence of mechanisms which buffer the observable profile against variations in the strength of the source. This strongly suggests that the mother can place a reproducible number of mRNA molecules into the egg, and that the apparatus for translation has an efficiency that is constant from embryo to embryo as well. It would be attractive to have direct measurements that confirm these conclusions. Of course, this also pushes the problem back. How does the mother count mRNA molecules with $\sim 10\%$ accuracy? How does the embryo ensure that the efficiency of translation, which depends on myriad factors, is reproducible?

Can we make the same argument in any other system? Maybe the Dpp experiments of Bollenbach et al (2008)? Others?

[71] It is not so easy to interpret these results quantitatively, because we don't really know if adding more copies of the gene produces *proportionately* higher concentrations of the protein. Still, Fig 120 is prima facie evidence against robustness of the pattern to variations in the strength of the source.

The problem we have been discussing thus far emerges as soon as we claim that position in the embryo is encoded by the concentration of specific molecules. In such a scheme, if we want neighboring cells to do different things, reliably, then we will be driven to questions about how these cells can distinguish small differences in concentration, as discussed in Section II.B.[72] Conversely, if we want two cells that occupy corresponding positions in different embryos to do the same thing, then we are driven to ask how the concentrations at these corresponding points can be the same. These issues of precision and reproducibility arise even if the size of the embryo and the external conditions of development are identical. There is another problem, related to the variations in size of the embryo, and this is the problem of scaling.

To a remarkable extent, the *proportions* of organisms are constant, despite size variations. We all know people who have especially large heads, but certainly the proportions of the body vary much less than the overall size, and again insects provide clear examples of this, both within species and across species. Different species of flies, for example, have embryos that span a factor of five or more in length, yet they have the same number of body segments, and individual segments have dimensions that scale with the overall size of the organism. You can see this scaling not just in the macroscopic patterns of the developed organisms, but also in the patterns of gene expression, as shown in Fig 122. Indeed, when we have looked at the problem of reproducibility above, we have implicitly used the idea of scaling, always plotting position as a fractional distance along the anterior–posterior axis.[73]

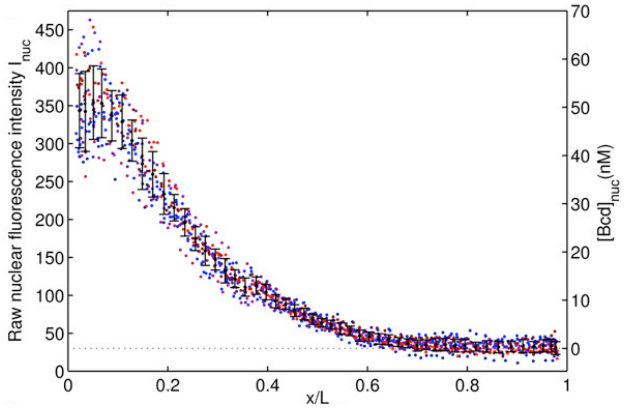

FIG. 121 Measurements of the Bicoid concentration in nuclei along the anterior–posterior axis of the *Drosophila* embryo, from Gregor et al (2007b). Each point corresponds to one nucleus in one embryo; points of the same color come from the same embryo, and error bars show the means and standard deviations across the fifteen embryos in the experiment. The vertical axis shows the fluorescence signal in embryos engineered to make the Bcd–GFP fusion protein, which can be calibrated to give the absolute concentration (at left). The horizontal axis shows the position of the nucleus as a fraction of the overall length of the embryo.

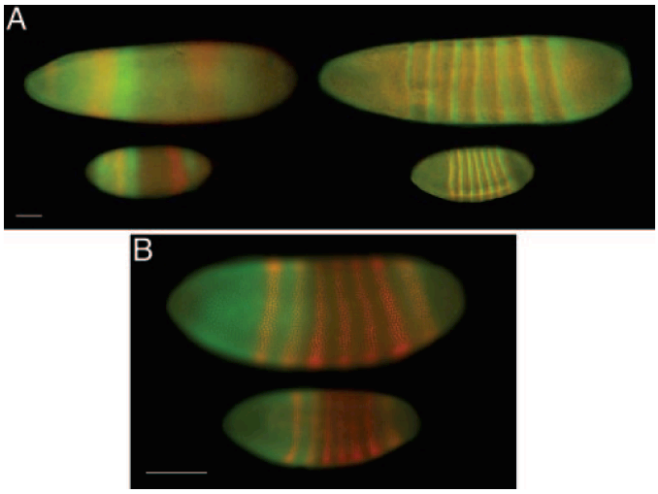

FIG. 122 Immunofluorescence stainings for products of the gap and pair–rule genes in flies of different sizes, from Gregor et al (2005). (A) Staining of *L sericata* (upper embryos) and *D melanogaster* (lower embryos) for Hunchback (green) and Giant (red) in the left column, and for Paired (green) and Runt (red) in the right column. (B) Staining of *D melanogaster* (upper embryos) and *D busckii* (lower embryo) for Hunchback (green) and Runt (red). Scale bars: 100 μm. [Should give typical sizes of the embryos in the different species!]

Scaling is deeply puzzling, perhaps more so for physicists who have thought about pattern formation in non–biological systems. To make this point, let's imagine making a model of the whole network of interactions that lead to, for example, the beautiful stripes of gene expression. In each nucleus there are chemical reactions corresponding to the transcription of the relevant genes, and the rates of these reactions are determined by the concentrations of the appropriate transcription factors. More equations will be needed to describe translation (although maybe one can simplify, if, for example, mRNA molecules degrade quickly and proteins live longer). Different points in space are coupled, presumably through diffusion of all these molecules, although we should worry

[72] See also the discussion of positional information, in bits, in Section IV.A.

[73] [I want to emphasize the distinction between the problems of reproducibility and scaling, but need to think about how to do this. For example, in Fig 121 the embryos have lengths with standard deviation of only $\sim 4\%$.]

about whether diffusion is the correct description. Even if youre not sure about the details, you can see the *form* of the equations: some sort of partial differential equations, in which the local time dependence of concentrations has contributions both from nonlinear terms describing the various chemical reactions and from spatial derivatives describing diffusion or other transport processes,

$$\frac{\partial g_i(x,t)}{\partial t} = D_i g_i(x,t) + F_i(\{g_j\}). \tag{570}$$

But we have seen equations like these before in the study of non–biological pattern forming systems such as Rayleigh–Bernard convection, directional solidification, ...

Many non–biological pattern forming system generate periodic spatial patterns that remind us of the segments in the insect and the patterns of pair rule gene expression. The scale of these patterns, however, is set by combinations of parameters in the equations. For example, we can combine a diffusion constant with a reaction rate or lifetime to get a length, as in the discussion of the Bicoid profile above ($\lambda = \sqrt{D\tau}$). What happens if you put these equations in a larger box? Well, from Rayleigh–Bernard convection, we know the answer. [should really have an image of convection, or some other 'physical' pattern formation problem]. In this system—a fluid layer heated from below—we see a collection of convective rolls, sometimes in stripes and sometimes in 2D cellular patterns. Again, the length scale of the stripes is determined by the parameters of the equation(s). If you put the whole system in a bigger box, you get more stripes, not wider stripes.

Problem 118: A lightning review of pattern formation. [give the students a tour of instabilities etc in some simple case!]

The results in Fig 122 come close to saying that we can put all the same equations into a bigger box, and the stripes come out wider in proportion to the length of the box. One might worry that these are different organisms, and so perhaps evolution has tuned the properties of the proteins involved so that the relevant combinations of parameters turn out to scale with embryo size. The differences can't be too large, because we can identify the same molecules as being involved through similarities of amino acid sequence, and because the same antibodies react with these molecules in different species. Still, it is possible that scaling across embryos in different species reflects an evolutionary adaptation.

If we look across related species of flies with embryos of very different sizes, then the Bcd profiles (as measured with antibody staining) seem to scale with the length of the egg. One can use the same experimental methods used in making the Bcd–GFP fusion more aggressively, extracting the sequences of Bicoid from flies of different sizes and re–inserting green versions of these different Bicoids into the *Drosophila* genome. The striking result is that the resulting spatial profiles are those appropriate to the host embryo, not the source of the Bicoid. Taken together, all of these results suggest that, as with the problem of variability, the scaling problem is solved at the level of Bicoid itself. It would appear that there is something about the environment or geometry of the embryo itself that couples the global changes in the size of the embryo to the local dynamics.

Scaling might not be so mysterious. Suppose that we think of the (roughly ellipsoidal) embryo as a cylinder, with the source covering one end of this cylinder; since most of the interior of the egg is yolk, we imagine that all degradation of proteins occurs near the surface. If the degradation reaction is rapid, then the surface of the embryo acts as a sink, and in the interior of the embryo the concentration obeys the diffusion equation, with no additional terms. Assuming cylindrical symmetry, the steady state profile must then obey

$$\frac{\partial^2 c_s(x,r)}{\partial x^2} + \frac{1}{r}\frac{\partial}{\partial r}\left[r\frac{\partial c_s(x,r)}{\partial r}\right] = 0 \tag{571}$$

$$-D\frac{\partial c_s(x,r)}{\partial x} = \frac{R}{\pi r_0^2} \tag{572}$$

$$c(x, r = r_0) = 0, \tag{573}$$

where x measures position along the axis of the cylinder (the anterior–posterior axis of the embryo), r_0 is the radius of the cylinder, R is once again the number of molecules per second being injected by the source, and the last condition follows in the limit that degradation reactions at the surface are fast. If we using the standard separation of variables method and look for solutions of the form $c_s(x,r) = e^{-x/\lambda} f(r)$, then we must have

$$\frac{1}{r}\frac{d}{dr}\left[r\frac{df(r)}{dr}\right] + \frac{1}{\lambda^2} f(r) = 0, \tag{574}$$

with the boundary condition $f(r_0) = 0$. You may recognize this as the differential equation which defines the Bessel function,

$$\frac{d^2 J_0(r)}{dr^2} + \frac{1}{r}\frac{dJ_0(r)}{dr} + J_0(r) = 0, \tag{575}$$

so that $f(r) \propto J_0(r/\lambda)$. But then to obey the boundary condition at the surface of the cylinder, we must have $\lambda = r_0/z_{01}$, where z_{01} is the location of the first point where $J_0(z) = 0$. So, in this model, the length scale of the Bicoid profile λ is automatically proportional to the radius of the embryo; if variations in aspect ratio are smaller than variations in length, this will serve, at least approximately, to scale the profile to the size of the egg.

Problem 119: Could it be so simple?. Work out the details of the scenario in the last paragraph. Looking at images of the fly embryo earlier in this section, estimate the radius r_0 assuming that the length of the embryo is $L \sim 0.5\,\text{mm}$. Does the prediction $\lambda = r_0/z_{01}$ actually work quantitatively?

While simple geometrical mechanisms of scaling might be too simple to work, we should note that embryos of different linear dimensions have the same number of cells. Further, because the nuclei arrange themselves more or less regularly over the embryo surface, the distance from one nucleus to the next provides a local measure proportional to the global size of the egg. Finish this discussion!!

Complementary to the problem of scaling is the problem of size control. In many developmental problems (even in later fly development), tissues are growing as they differentiate, and cells have to know both where they are and whether they should still be dividing and hence expanding the size of the tissue. [Add discussion of work by Shraiman on size control, and subsequent experiments.]

The discussion so far has taken very seriously the idea that there are "primary morphogens," placed by the mother, which define provide the basic signal for position in the embryo. Position is a continuous variable, as is concentration. A very different perspective emphasizes that, when development is finished, cells have adopted distinct types or "fates," which define their function in the adult organism. These fates persist long after the primary morphogen signals have disappeared, and so they must represent stable states of the cells, thus bringing us back to the theme of this section. Cells even maintain their identity and state when separated from their neighbors, which suggests that the biochemical and genetic networks in each cell have multiple attractors. A minimal model of the networks relevant for development, then, would have the right number of attractors but a limited number of dynamical variables, perhaps much fewer than the number of genes involved in the entire network. As with the attractors in the Hopfield model, there is a plausible path to "robustness," because changing the qualitative behavior of the system would actually require changing the number of attractors—the development of cells into types becomes a matter of topology rather than geometry in the model, and hence invariant to a finite range of parameter variation.

Need to fill out the discussion of attractors. In some ways this is a mathematization of Waddington's "canalization," which is an old idea. In modern times, there is work by Reinitz, Sharp and colleagues that tries to make a more direct analogy between genetic and neural networks. Most recently there is work by Siggia and Carlson on vulva development in *C elegans* that pushes the "minimal model" strategy the furthest, arguing that we can choose coordinates to make the attractors obvious, and then try to map the known biochemical signals into these coordinates, rather than the more usual effort to use biochemical coordinates and decipher the attractors. This belongs here, but isn't published yet .. hopefully by the time I finalize the text there will be something to cite.

This section needs a conclusion. We have covered a lot of territory, from chemotaxis to development ... what have we learned?

Some of the basic idea about adaptation in sensory neurons were established early on, by Adrian and Zotterman; for a review see Rieke et al (1997).

Adrian 1926: The impulses produced by sensory nerve endings: Part I. ED Adrian, *J Physiol (Lond)* **61,** 49–72 (1926).

Adrian & Zotterman 1926a: ED Adrian & Y Zotterman, The impulses produced by sensory nerve endings: Part II. The response of a single end organ. *J Physiol (Lond)* **61,** 151–171 (1926).

Adrian & Zotterman 1926b: ED Adrian & Y Zotterman, The impulses produced by sensory nerve endings: Part III. Impulses set up by touch and pressure. *J Physiol (Lond)* **61,** 465–483 (1926).

Rieke et al 1997: *Spikes: Exploring the Neural Code.* F Rieke, D Warland, R de Ruyter van Steveninck & W Bialek (MIT Press, Cambridge, 1997).

[Need to check on references for adaptation in bacterial chemotaxis in Chapter 2] Renewed interest in this system was triggered by the work of Barkai and Leibler (1997), who used adaptation in chemotaxis as an example for the more general problem of robustness. The idea that perfect adaptation could be achieved even in the presence of variations in protein copy numbers was then tested more directly by Alon et al (1999). [Need to reference subsequent work that goes beyond the mean–field level, e.g. from Wingreen et al] Recent work suggests that, although the biochemical network responsible for chemotaxis may allow for robustness against variations in protein copy numbers, under natural conditions there is relatively precise control over (at least) relative copy numbers, even for proteins on different operons (Kollman et al 2005). In competition experiments, one can even show that tight correlations between protein concentrations improves chemotactic performance (Løvdok et al 2009).

Alon et al 1999: Robustness in bacterial chemotaxis. U Alon, MG Surette, N Barkai & S Leibler, *Nature* **397,** 168–171 (1999).

Barkai & Leibler 1997: Robustness in simple biochemical networks. N Barkai & S Leibler, *Nature* **387,** 913–917 (1997).

Kollman et al 2005: Design principles of a bacterial signalling network. M Kollmann, L Løvdok, K Bartholome, J Timmer & V Sourjik *Nature* **438,** 504–507 (2005).

Løvdok et al 2009: Role of translational coupling in robustness of bacterial chemotaxis pathway. L Løvdok, K Bentele, N Vladimirov, A Müller, FS Pop, D Lebiedz, M Kollmann & V Sourjik, *PLoS Biology* **7,** e1000171 (2009).

Pointers toward work on the cell cycle.

Bean et al 2006: Coherence and timing of cell cycle start examined at single–cell resolution. JM Bean, ED Siggia & FR Cross, *Mol Cell* **21,** 3–14 (2006).

Di Talia et al 2007: The effects of molecular noise and size control on variability in the budding yeast cell cycle. S Di Talia, JM Skotheim, JM Bean, ED Siggia & FR Cross, *Nature* **448,** 947–952 (2007).

Li et al 2004: The yeast cell cycle network is robustly designed. F Li, Y Lu, T Long, Q Ouyang & C Tang, *Proc Nat'l Acad Sci (USA)* **101,** 4781 (2004).

Zhang et al 2006: A stochastic model of the yeast cell cycle network. Y Zhang, M Qian, Q Ouyang, M Deng, F Li & C Tang, *Physica D* **219,** 35 (2006).

Lau et al 2007: Function constrains network architecture and dynamics: A case study on the yeast cell cycle Boolean network. K Lau, S Ganguli & C Tang, *Phys Rev E* **75,** 051907 (2007).

Skotheim et al 2008: Positive feedback of G1 cyclins ensures coherent cell cycle entry. JM Skotheim, S Di Talia, ED Siggia & FR Cross, *Nature* **454,** 291–297 (2008).

A modern textbook account of development in the fly embryo is provided by Lawrence (1992). We know which genes are relevant to the earliest events in patterning because of pioneering experiments first by EB Lewis and then by EF Wieschaus and C Nüsslein–Vollhard. Lewis identified a series of puzzling mutant flies where a mutation in a single gene could generate flies that were missing segments, or had extra segments. It is as if the "program" of embryonic development has subroutines (!). Wieschaus and Nüsslein–Vollhard decided to search for all the genes such that mutations in those genes would perturb the formation of spatial structure in the embryo, and they found that there are surprisingly few such genes, on the order of 100. To get a feeling for all this, one can certainly do worse than to read the Nobel lectures from 1994 (Lewis 1997; Nüsslein–Volhard 1997; Wieschaus 1997).

Lawrence 1992: *The Making of a Fly: The Genetics of Animal Design* PA Lawrence (Blackwell, Oxford, 1992).

Lewis 1995: The bithorax complex: The first fifty years. EB Lewis, in *Nobel Lectures, Medicine or Physiology 1991–1995* N Ringertz, ed, pp 247–272 (World Scientific, Singapore, 1997).

Nüsslein–Volhard 1997: The identification of genes controlling development in flies and fishes. C Nüsslein–Volhard, in *Nobel Lectures, Medicine or Physiology 1991–1995* N Ringertz, ed, pp 285–306 (World Scientific, Singapore, 1997).

Wieschaus 1997: Molecular patterns to morphogenesis: The lessons from *Drosophila*. EF Wieschaus, in *Nobel Lectures, Medicine or Physiology 1991–1995* N Ringertz, ed, pp 314–326 (World Scientific, Singapore, 1997).

The classical ideas about pattern formation in non–equilibrium systems were presented by Turing (1952), who was aiming specifically at an understanding of embryonic development. Modern views are given by Cross & Hohenberg (1993) and by Cross & Greenside (2009).

Cross & Greenside 2009: *Pattern Formation and Dynamics in Nonequilibrium Systems* M Cross & H Greenside (Cambridge University Press, Cambridge 2009).

Cross & Hohenberg 1993: Pattern formation outside of equilibrium. MC Cross & PC Hohenberg, *Revs Mod Phys* **65,** 851–1112 (1993).

Turing 1952: The chemical basis of morphogenesis. AM Turing, *Phil Trans R Soc Lond B* **237,** 33–72 (1952).

The general idea that cells know their position, and hence their fate, in an embryo by responding to the concentration of some special "morphogen" molecule is very old, and it didn't take too long before people started to think about the role of diffusion in establishing morphogen gradients. Some milestones are Wolpert's discussion of positional information (Wolpert 1969), and Crick's surprisingly influential discussion of diffusion (Crick 1970). The transcription factor bicoid, in the *Drosophila* embryo, provides a very clear example of these ideas (Driever & Nüsslein–Vollhard 1988a,b; Ephrussi & St Johnston 2004). I am embarrassed not to know who first wrote down the simple model for Bcd profiles, and I should check!

Crick 1970: Diffusion in embryogenesis. F Crick, *Nature* **225,** 420–422 (1970).

Driever & Nüsslein–Vollhard 1988a: A gradient of Bicoid protein in *Drosophila* embryos. W Driever & C Nüsslein–Vollhard, *Cell* **54,** 83–93 (1988).

Driever & Nüsslein–Vollhard 1988b: The Bicoid protein determines position in the *Drosophila* embryo in a concentration–dependent manner. W Driever & C Nüsslein–Vollhard, *Cell* **54,** 95–104 (1988).

Ephrussi & St Johnston 2004: Seeing is believing: The bicoid morphogen gradient matures. A Ephrussi & D St Johnston, *Cell* **116,** 143–152 (2004).

Wolpert 1969: Positional information and the spatial pattern of cellular differentiation. L Wolpert, *J Theor Biol* **25,** 1–47 (1969).

Houchhmandzadeh et al (2002) drew attention to the problem of variability in morphogen gradients, and their suggestion that the emergence of reproducible patterns was an example of robustness in biochemical networks attracted considerable attention. Among the models that emerged in an attempt to flesh out the idea of robustness, some make specific use of gradients from the two ends of the embryo to compensate for global parameter variations and allow for scaling with the size of the egg (Houchmandzadeh et al 2005, McHale et al 2006), while others use nonlinearities in degradation reactions (Eldar et al 2002) or in the transport process (Bollenbach et al 2005) to generate spatial profiles that are robust against variations in source strength. Although much of this discussion focuses on early events in embryonic development, there is also the idea that the final patterns of gene expression, which are more closely tied to cell fate, should be robust steady states of the relevant biochemical networks (von Dassow et al 2000). Even earlier work emphasized the similarity of these networks to neural nets, with stable patterns being analogous to stored memories (Mjolsness et al 1991), and one can see this as a modern formulation of the ideas of "canalization" (Waddington 1942). Most recent work from Siggia & Carlson. Have to see what gets said about size control, but certainly will cite Shraiman (2005).

Bollenbach et al 2005: Robust formation of morphogen gradients. T Bollenbach, K Kruse, P Pantazis, M Gonzalés–Gaitán & F Jülicher, *Phys Rev Lett* **94,** 018103 (2005).

von Dassow et al 2000: The segment polarity network is a robust developmental module. G von Dassow, E Meir, EM Munro & GM Odell, *Nature* **406,** 188–192 (2000).

Eldar et al 2002: Robustness of the BMP morphogen gradient in *Drosophila* embryonic development. A Eldar, R Dorfman, D Weiss, H Ashe, B–Z Shilo & N Barkai, *Nature* **419,** 304–308 (2002).

Houchmandzadeh et al 2002: Establishment of developmental precision and proportions in the early *Drosophila* embryo. B Houchmandzadeh, E Wieschaus & S Leibler, *Nature* **415,** 798–802 (2002).

Houchmandzadeh et al 2005: Precise domain specification in the developing *Drosophila* embryo. B Houchmandzadeh, E Wieschaus & S Leibler, *Phys Rev E* **72,** 061920 (2005).

McHale et al 2006: Embryonic pattern scaling achieved by oppositely directed morphogen gradients. P McHale, W–J Rappel & H Levine, *Phys Biol* **3 ,** 107–120 (2006).

Mjolsness et al 1991: A connectionist model of development. E Mjolsness, DH Sharp & J Reinitz, *J Theor Biol* **152,** 429–453 (1991).

Shraiman 2005: Mechanical feedback as a possible regulator of tissue growth. BI Shraiman, *Proc Nat'l Acad Sci (USA)* **102,** 3318–3323 (2005).

Waddington 1942: Canalization of development and the inheritance of acquired characters. CH Waddington, *Nature* **150,** 563–565 (1942).

For measurements on the reproducibility of the early events in the fly embryo, see Dubuis et al (2011). Are there classical references? As noted above, and in Section 2.3, the overall precision and reproducibility of pattern formation in the fruit fly embryo is equivalent to $\sim 10\%$ accuracy in the concentration of Bicoid. Although this might not be how things actually work, it does suggest a standard for making measurements of the Bicoid concentration (and, perhaps, for other morphogens as well). For a recent discussion of the state of the art in these experiments, see Dubuis et al (2010). The measurements on reproducibility of the Bcd profiles shown in Fig 121 are from Gregor et al (2007b), cited in Section [** General decision—is it ok to give references more than once in different sections?]. Experiments on the scaling of Bcd profile across species include Gregor et al (2005) and Gregor et al (2008).

Dubuis et al 2010: Quantifying the Bicoid morphogen gradient in living fly embryos. J Dubuis, AH Morrison, M Scheeler & T Gregor, arXiv:1003.5572 [q–bio.QM] (2010).

Dubuis et al 2011: Positional information, in bits. JO Dubuis, G Tkačik, W Bialek, EF Wieschaus & T Gregor, in preparation (2011).

Gregor et al 2005: Diffusion and scaling during early embryonic pattern formation. T Gregor, W Bialek, DW Tank, RR de Ruyter van Steveninck, DW Tank & EF Wieschaus, *Proc Nat'l Acad Sci (USA)* **102,** 18403–18407 (2005).

Gregor et al 2008: Shape and function of the Bicoid morphogen gradient in dipteran species with different sized embryos. T Gregor, AP McGregor & EF Wieschaus, *Dev Biol* **316,** 350–358 (2008).

D. Long time scales in neural networks

The basic time scales of electrical dynamics in neurons are measured in milliseconds, yet the time scales of our mental experience are much longer. From the fraction of a second that we need to integrate sounds as we identify words or phrases, to the minutes of memory for a phone number, to the decades over which our recollections of childhood experiences can stretch, the brain has access to time scales far beyond those describing the elementary events of action potential generation and synaptic transmission. If we write a set of dynamical equations, and the time scales which emerge to describe the whole system are much longer than the time scales which appear as parameters in the equations, then something special has happened. How does this work in the brain? How does the system insure that this seemingly special separation of time scales occurs robustly?

One possible solution to the wide range of relevant time scales is to invoke a correspondingly wide range of mechanisms, and surely this is part of the right answer. Thus, it seems unlikely that memories of things long past are stored as continuing patterns of electrical activity in the brain, which somehow last for $\sim 10^{10}\times$ longer than their natural time scale, and are always present to be examined as we reminisce. On the other hand, for working memory—holding the words of a sentence in our minds, or doing mental arithmetic—the time scales involved seem at once long compared with natural time scales for electrical activity, yet too short to engage biochemical mechanisms, such as the regulation of gene expression, which could have more stable, semi–permanent effects.

In fact, we know a whole class of examples in which long time scales emerge naturally. When a ball rolls down a hill, the time scale of the rolling may be short, but once at the bottom the ball can stay there (more or less) forever. So, perhaps we can arrange for the dynamics of neurons in an interconnected network to be like the motion of a particle on a (multidimensional) landscape, with nice deep valleys corresponding to patterns of activity that can persist for a long time once the system find itself in the right neighborhood. In two hugely influential papers in the early 1980s, Hopfield showed how to do exactly this.

A typical neuron in the brain receives inputs from many other neurons [need to see where we've had a chance to talk about axons, dendrites, synapses .. should be before this!]; in the cortex 'many' is several thousand, and in the extreme case of the cerebellum 'many' actually means $\sim 10^5$. Conversely, although each cell has only one axon along which its output action potentials are sent, this axon can branch to contact thousands of

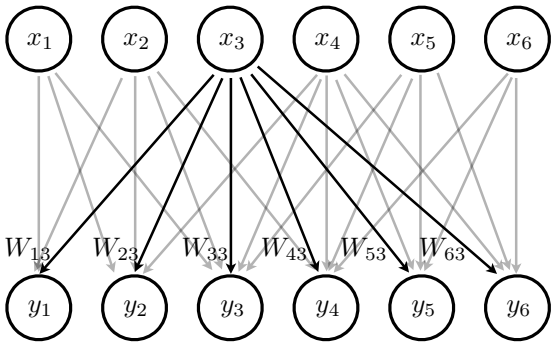

FIG. 123 A schematic network of neurons, focusing on one cell i that receives inputs from may other cells j = 1, 4, 8, $\cdots$.

other cells. Let's focus on one cell i, which receives inputs from many other cells j, as in Fig 156. Schematically, we can imagine that each cell is either active or inactive, on or off, and hence the state of one cell can be represented by a binary variable $\sigma_i = \pm 1$; for the moment we will leave this as schematic, and not try to interpret σ_i too closely in terms of action potentials or membrane voltage. In the simplest view, each cell j sends its output to cell i, and as these inputs are collected from the synapses, they are summed with some weights W_{ij} which we can think of as the "strengths" of the synapses from cell j to cell i. Having summed its inputs, cell i must then decide whether to be on or off, comparing the total input to a threshold θ_i. These words are equivalent to saying that the state of cell i is set according to the equation

$$\sigma_i \to \mathrm{sgn}\left[\sum_j W_{ij}\sigma_j - \theta_i\right]. \tag{576}$$

Models of this flavor go back at least to the 1940s, when McCulloch and Pitts explored the idea that the on/off states of neurons could implement a kind of logical calculus. Precisely because they can perform such operations, these sorts of discrete dynamics can be almost arbitrarily complicated. Thus, in general, it's hard to say anything about the dynamics generated by Eq (576).

Suppose, however, that if neuron j synapses onto neuron i with strength W_{ij}, then neuron i synapses onto neuron j with the same strength, so that the matrix of synaptic strengths W_{ij} is symmetric. Then the updating of the state of neruon i in Eq (576) serves to reduce an 'energy' function defined by

$$E = -\frac{1}{2}\sum_{ij}\sigma_i W_{ij}\sigma_j + \sum_i \theta_i\sigma_i. \tag{577}$$

Indeed, we recognize Eq (576) as being the dynamics of a zero temperature Monte Carlo simulation of an Ising model with energy defined by Eq (577). Now, we can make progress.

Problem 120: Energy in the Hopfield model. Show explicitly that the dynamics in Eq (576) serves to decrease the energy function in Eq (577).

If we can map the dynamics of a neural network onto the Ising model, then we can bring an enormous amount of our intuition (and mathematical tools) from statistical mechanics. We know that, since the dynamics we have defined are at zero temperature—we are neglecting, for the moment, any noise in the neurons or synapses—it is possible to have collective states of the whole system which are stable forever. The simplest example is with all thresholds equal to zero, and all synaptic strengths equal and positive. Then the energy function becomes

$$E = -\frac{W}{2}\sum_{ij}\sigma_i\sigma_j = -\frac{W}{2}\left(\sum_i \sigma_i\right)^2. \tag{578}$$

This is the mean–field ferromagnet. In this model there are two stable ground states—all neurons 'on' ($\sigma_i = +1$ for all i) and all neurons 'off' ($\sigma_i = -1$ for all i). Two states aren't many, and these states seem especially odd, but maybe we are on the right track.

If instead of making all the W_{ij} equal, we choose them at random, then the Ising model we have constructed is the mean–field or Sherrington–Kirkpatrick spin glass. We know that this system has many locally stable states, with an energy landscape that has valleys within valleys, as discussed in Section III.A. This is probably too much, since the structure of these exponentially large number of states depends very sensitively on the precise form of the couplings W_{ij}. More generally, since we only have $\sim N^2$ parameters at our disposal when we adjust the W_{ij}, it is difficult to imagine how we could 'program' the network to store exponentially many independent patterns.

To find a compromise between the ferromagnet and the spin glass, we recall a trick from the history of models for magnetism. Suppose that

$$W_{ij} = W\xi_i\xi_j, \tag{579}$$

where $\vec{\xi}$ is an arbitrary binary vector, $\xi_i = \pm 1$, and for simplicity let the thresholds $\theta_i = 0$. Then the energy becomes

$$\begin{aligned} E &= -\frac{1}{2}\sum_{ij}\sigma_i W_{ij}\sigma_j \\ &= -\frac{W}{2}\sum_{ij}\sigma_i\xi_i\xi_j\sigma_j \qquad (580)\\ &= -\frac{W}{2}\sum_{ij}(\xi_i\sigma_i)(\xi_j\sigma_j) \qquad (581)\\ &= -\frac{W}{2}\sum_{ij}\tilde{\sigma}_i\tilde{\sigma}_j, \qquad (582)\end{aligned}$$

where $\tilde{\sigma}_i = \xi_i\sigma_i$ is again a binary variable, $\tilde{\sigma}_i = \pm 1$. The transformation $\sigma_i \to \tilde{\sigma}_i$ is a discrete gauge transformation, so we see that the model with weights in Eq (579) is gauge equivalent to a ferromagnet. Rather than the stable states of the system being $\sigma_i = +1$ for all i and $\sigma_i = -1$ for all i, the stable states are $\sigma_i = +\xi_i$ and $\sigma_i = -\xi_i$. Importantly, this construction can be generalized.

Rather than Eq (579), let us imagine that

$$W_{ij} = W\left(\xi_i^{(1)}\xi_j^{(1)} + \xi_i^{(2)}\xi_j^{(2)}\right). \tag{583}$$

Now we have

$$E = -\frac{1}{2}\sum_{ij}\sigma_i W_{ij}\sigma_j$$
$$= -\frac{W}{2}\left[\sum_{ij}\sigma_i\xi_i^{(1)}\xi_j^{(1)}\sigma_j\right] - \frac{W}{2}\left[\sum_{ij}\sigma_i\xi_i^{(2)}\xi_j^{(2)}\sigma_j\right] \quad (584)$$
$$= -\frac{W}{2}\left[\left(\vec{\xi}^{(1)}\cdot\vec{\sigma}\right)^2 + \left(\vec{\xi}^{(2)}\cdot\vec{\sigma}\right)^2\right]. \quad (585)$$

Clearly the energy will be low if the pattern of neural activity $\vec{\sigma}$ is parallel to the vector $\vec{\xi}^{(1)}$ *or* to the vector $\vec{\xi}^{(2)}$. But in a high dimensional space, two randomly chosen vectors are, with high probability, nearly orthogonal. This means that the two terms in the Hamiltonian can't both be important at once. Thus, the energy function will have a minimum near $\vec{\sigma} = \vec{\xi}^{(1)}$ and a separate minimum near $\vec{\sigma} = \vec{\xi}^{(2)}$, as well as the flipped versions of these states, $\vec{\sigma} = -\vec{\xi}^{(1)}$ and $\vec{\sigma} = -\vec{\xi}^{(2)}$.

Problem 121: Random vectors in high dimensions. Consider random binary vectors $\vec{v}$ in an N–dimensional space: $\vec{v} \equiv \{v_1, v_2, \cdots, v_N\}$, where each $v_i = \pm 1$ is chosen independently and at random. The angle ϕ between two such vectors is defined in the usual way by normalizing the dot product,

$$\cos\phi \equiv \frac{1}{N}\vec{v}^{(1)}\cdot\vec{v}^{(2)}. \quad (586)$$

Before calculating anything, explain why, if $\vec{v}^{(1)}$ and $\vec{v}^{(2)}$ are chosen independently, it must be that $\langle\cos\phi\rangle = 0$. Calculate the variance $\langle\cos^2\phi\rangle$ to show that the typical values of $|\cos\phi| \sim 1/\sqrt{N}$, which vanishes as $N \to \infty$. Can you use the central limit theorem to say something about the whole probability distribution $P(\cos\phi)$ in this limit? Show that the distribution can be written exactly as

$$P(z = \cos\phi) = \int \frac{dk}{2\pi} e^{-ikz}\left[\cos(k/N)\right]^N. \quad (587)$$

Connect this result to the predictions of the central limit theorem. Develop a saddle point approximation so that you can calculate, at large N, $P(z)$ for values of $|z| \gg 1/\sqrt{N}$. Verify your approximations with a simulation.

The key idea now is to go further, with not just two patterns but many, writing the weights as

$$W_{ij} = W\sum_{\mu=1}^{K}\xi_i^{(\mu)}\xi_j^{(\mu)}. \quad (588)$$

Then the energy becomes

$$E = -\frac{W}{2}\sum_{ij}\sigma_i\left[\sum_{\mu=1}^{K}\xi_i^{(\mu)}\xi_j^{(\mu)}\right]\sigma_j = -\frac{W}{2}\sum_{\mu=1}^{K}\left(\vec{\xi}^{(\mu)}\cdot\vec{\sigma}\right)^2. \quad (589)$$

Certainly if $K \ll N$ our intuition from the case of two patterns should carry over, since almost all of the vectors $\vec{\xi}^{(\mu)}$ will be nearly orthogonal, and we should find that the energy function has $2K$ minima, near the vectors $\pm\vec{\xi}^{(\mu)}$. At some value of K this must stop being true; indeed if we let K itself become large we must get back to the spin glass model in which there are many locally stable states, but they don't have any connection to the patterns $\vec{\xi}^{(\mu)}$ that we have 'programmed' into the system. In his original work on this model, Hopfield gave rough arguments to suggest that this transition from ordered to disordered behavior occurs at roughly $K \sim 0.15N$, so that it should be possible to have a number of states which is proportional to the number of neurons, and he verified this in simulations with $N = 100$ [should break this off as a paragraph and give the argument, rather than pointing].

Problem 122: Simulating the Hopfield model. Given a matrix W_{ij} it is straightforward to simulate the dynamics of the Hopfield model, as defined by Eq (576); try the simplest case, with $\theta_i = 0$. To run the simulation, you can go through these steps:

1. Start a collection of N spins in some randomly chosen state.
2. Choose one spin i at random.
3. Set $\sigma_i = \mathrm{sgn}\left[\sum_j W_{ij}\sigma_j\right]$.
4. Choose another spin and repeat the update, again and again

Produce a series of simulations to convince yourself that, with W_{ij} chosen as in Eq (588) and a small value of K, the dynamics always stop in the neighborhood of one of the vectors $\vec{\xi}^{(\mu)}$ that you have used in sculpting the energy landscape. Explore what happens as K becomes larger. If you jump to $K \sim N/2$, can you see the emergence of more random stopping points for the dynamics? Perhaps even if you start at one of the vectors $\vec{\xi}^{(\mu)}$, the interference from the other vectors destabilizes this state? If the dynamics stops at a state $\vec{\sigma}_s$, define an order parameter by finding the nearest vector $\vec{\xi}^{(\mu)}$, and measuring the normalized dot product,

$$m_s = \max_\mu \left|\vec{\xi}^{(\mu)}\cdot\vec{\sigma}_s\right|. \quad (590)$$

From many random starting points, what is the mean value of m_s as a function of K and N? As N gets larger, do you see the emergence of a 'thermodynamic limit,' where the (intensive) order parameter $\langle m_s\rangle$ depends only on the ratio K/N? Are there signs of a phase transition at some critical value of K/N?

The idea that the dynamics of neural networks could be mapped onto the Ising model immediately captured the imagination of the physics community. But before exercising ourselves in this direction, let's think about how much progress we have made toward solving our original problem. The Hopfield model shows how the dynamics of a neural network can correspond to 'downhill' motion on an energy landscape, much like a ball rolling down a

hill. Thus, the system as a whole has collective, macroscopic states which will persist for times arbitrarily long compared with the basic time scales of the system, the time scales on which the individual neurons update their microscopic states according to Eq (576). Importantly, there are not just a few of these stable states, but many, in proportion to the number of neurons. Unlike the case of the ball coming to a stop at the bottom of the hill, the stability of these states is the result of activity, each neuron receiving continuous input from other neurons in the network; in effect the stable states are patterns of electrical activity which can reinforce themselves as they propagate through the network, embodying old ideas about the 'reverberation' of activity patterns through the extensive feedback loops found in the brain.

It is tempting to think of the stable patterns of activity, $\vec{\sigma} \approx \vec{\xi}^{(\mu)}$ as *being* memories. When we set the synaptic connection matrix to the form shown in Eq (588), we "store" the memories, and as the dynamics settles into one of its locally stable states, one of these memories is "recalled." Each of the stored memories has a large basin of attraction, so the network will recall the memory given only a relatively weak "hint" that the memory is somewhere in the neighborhood of the current state. I use quotation marks extensively here to highlight the fact that we are sliding from properties of the equations into the everyday language that we use in describing our internal mental experiences, and this is dangerous. But, of course, it is also great fun.

A crucial property of the model is that a particular memory—e.g., $\mu = 42$—is not stored in any particular place. There is no single neuron or synapse that has responsibility for remembering this single recallable item. Instead, the memory is distributed over essentially all of the elements in the system. Correspondingly, if we eliminate one neuron or one synapse, there is no catastrophic loss of one memory, but at worst a gentle degradation of all the memories; in the limit $K \ll N$ we might even imagine that, as $N \to \infty$ deletion of anything less than a finite fraction of cells or synapses would have a vanishingly small effect. This 'fault tolerance' is a highly attractive property.

Problem 123: Fault tolerance. Develop a small simulation to illustrate the idea of fault tolerance in the Hopfield model.

One might worry that all of this depends upon a very particular form of the synaptic weight matrix, Eq (588). But this form is both natural and, perhaps surprisingly, well connected to experiment. Suppose that the current state of activity in the network, $\vec{\sigma}(t)$ represents something that we would like to store and be able to recall later. If every synaptic strength is changed by the rule

$$W_{ij} \to W_{ij} + W\sigma_i(t)\sigma_j(t), \tag{591}$$

then, assuming that we have not already tried to store too many patterns in the network, the current state $\vec{\sigma}(t)$ will act as one more pattern that can be recalled, one more stable state in the energy landscape—the network will have "learned" the state $\vec{\sigma}(t)$. Importantly, the change in strength of the synapse from neuron i to neuron j depends only on the states of neurons i and j. Thus, although the memory is distributed throughout the network, the rule for storing the memory is completely local.

The rule for modification of synaptic strengths in Eq (591), sometimes called a "learning rule," means that, over time, the strength of the synapse from neuron j to neuron i will be proportional to the correlation between the activities of these two cells. Learning based on correlations is an idea that goes back at least to Hebb in the 1940s, although there are clear precursors in the writing of William James fifty years earlier. Both James and Hebb were making an intuitive leap between the macroscopic phenomena of human and animal learning and what they imagined could be the underlying neural mechanisms. Although their words admit some breadth of interpretation, to a remarkable extent they were right, and many synapses are found to exhibit "Hebbian plasticity."

At this point we should say something about the experiments which demonstrate Hebbian plasticity at real synapses. Should get as far as explaining that there is a new issue of time scale separation, since the memory trace should be written quickly (so that the relevant biochemical mechanisms must switch quickly) but then be stable for long times, despite the fact that all the molecules get replaced fairly often. Models for this bring us back to the question of stability against noise in biochemical networks, which is something that should have been covered, in part in Section II.B. There is a lot that one could say here (one could make a nice course about synaptic plasticity alone), so careful selection is required.

What is the evidence that something like the Hopfield model is actually a correct description of real neural networks? The essence of the model, shorn of the analogies to magnetism, is that a recalled memory is a stable state of neural activity, one which persists in the absence of external stimuli by reverberating in the network. Persistent activity of neurons has been observed. The canonical example occurs when an animal has to remember a sensory stimulus for a brief time (a few second to a minute) in order to compare it with another image or more simply because an immediate response would be impossible. In the period between the stimulus and the cue for the response, where the subject has to remember what has been seen or heard, these neurons continue to generate action potentials at a rate very different from the 'resting' rate

before exposure to the initial stimulus, as shown in Fig 124. Although the behavior of each cell is different in detail, in many cases the activity during this 'delay period' is steady, as if the system were simply locked into a new state, but the state into which the system falls is different depending on the image which is being remembered. Persistent activity is not just a feature of our cortex, but appears also in many other systems, from the primate spinal cord to the goldfish brainstem. [Probably need more here: demonstrate that persistent activity varies in relation to the triggering inputs, in some cases is continuously graded, etc..]

One would like to demonstrate directly that the persistent activity of individual neurons during the delay period really reflects a collective state of the network. This is not so easy to do. Need to decide how far to go here—are there good experiments in cortex looking at synaptic inputs? Maybe say that this is an important reason to look for simpler examples ... Also want at least to point toward Amit's analysis of the Miyashita correlations, where the persistent patterns of activity have a trace of the sequence in which images were presented during learning.

At this point it would be nice to say a little about the more sophisticated analysis of the Hopfield model using replicas. The goal is to calculate the 'capacity,' that is the maximum number of patterns K that can be stored and successfully retrieved. This can be formulated as a problem in the statistical mechanics of disordered systems. I am not sure how much technical force is needed here (or in the discussion of protein above). Advice is welcome!

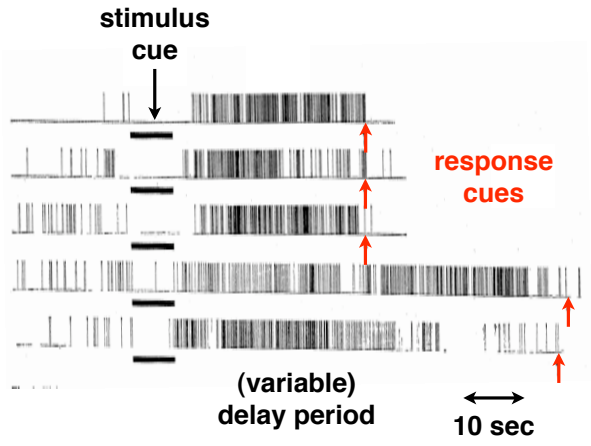

FIG. 124 The activity of a single neural in primate prefrontal cortex during short–term memory, from Fuster & Alexander (1971). In these experiments a rhesus monkey is trained to open one of two doors when he receives a cue that they are unlocked (response cues). Some time before this, the subject has been allowed to see which of the doors has a piece of apple behind it (stimulus cue). This neuron seems to be active during the delay period, and this persistent activity plausibly is part of the memory that the subject hold. These data record the results of five such experiments, where the vertical lines mark the times of spikes, and the arrows mark the times of the cues, as labelled.

There is a very different way of connecting the Hopfield model to experiment. Imagine that we divide time into small bins of duration $\Delta\tau$. If $\Delta\tau$ is sufficiently small, then each neuron either generate an action potential in this bin, or it does not not, so that the neural response is naturally binary: $\sigma_i = +1$ for a spike, $\sigma_i = -1$ for silence. for a large network it is impossible to 'measure' the probability distribution of all the network states, $P(\vec{\sigma})$. But even recording from neurons one by one it is possible to measure the mean rate at which each cell generates spikes, which is equivalent to the expectation value $\langle\sigma_i\rangle$, and it is becoming increasingly common to record at least from pairs of cells, which makes it possible to estimate the correlations $C_{ij} \equiv \langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle$. One could ask, as a purely practical question, what do these measurements tell us about the full distribution $P(\vec{\sigma})$? In general, of course, there are infinitely many distributions (over the 2^N states) that are consistent with these $N(N+1)/2$ measurements. Out of all these possible distributions, there is one which reproduces the measurements but otherwise describes a system which is as random or unstructured as possible, and this is the maximum entropy distribution, as we discussed in Section III.A; see also Appendix A.8.

We recall that the maximum entropy distribution consistent with a certain mean energy for a system is the Boltzmann distribution. This construction generalizes. Suppose that we are looking for the probability distribution $P(\vec{\sigma})$, and we know the expectation values of some functions on the state, $\langle f_\mu(\vec{\sigma})\rangle = \bar{f}_\mu$. Then to maximize the entropy of the distribution subject to these constraints, we use Lagrange multipliers as usual. Thus, our problem is to maximize [again, let's be careful about how this is done here vs. earlier vs. Appendix A.8]

$$\mathcal{F} = -\sum_{\vec{\sigma}} P(\vec{\sigma}) \ln P(\vec{\sigma}) - \sum_\mu \lambda_\mu \left[\sum_{\vec{\sigma}} P(\vec{\sigma}) f_\mu(\vec{\sigma}) - \bar{f}_\mu \right] - \Lambda \left[\sum_{\vec{\sigma}} P(\vec{\sigma}) - 1 \right], \quad (592)$$

where the last term fixes the normalization of the distribution. Following through the steps, the optimum is defined by

$$0 = \frac{\delta\mathcal{F}}{\delta P(\vec{\sigma})} = -\left[\ln P(\vec{\sigma}) + 1\right] - \sum_\mu \lambda_\mu f_\mu(\vec{\sigma}) - \Lambda \quad (593)$$

$$\ln P(\vec{\sigma}) = -\sum_\mu \lambda_\mu f_\mu(\vec{\sigma}) - (\Lambda + 1) \quad (594)$$

$$P(\vec{\sigma}) = \frac{1}{Z} \exp\left[-\sum_\mu \lambda_\mu f_\mu(\vec{\sigma}) \right], \quad (595)$$

where the partition function $Z = e^{-(\Lambda+1)}$, or, fixing normalization,

$$Z(\{\lambda_\mu\}) = \sum_{\vec{\sigma}} \exp\left[-\sum_\mu \lambda_\mu f_\mu(\vec{\sigma})\right]. \tag{596}$$

The multipliers λ_μ are determined by matching the expectation values in the distribution to those observed experimentally. We recall the usual identity

$$\langle f_\nu(\vec{\sigma})\rangle = -\frac{\partial \ln Z(\{\lambda_\mu\})}{\partial \lambda_\nu}, \tag{597}$$

so we have to solve the equations

$$-\frac{\partial \ln Z(\{\lambda_\mu\})}{\partial \lambda_\nu} = \bar{f}_\mu \tag{598}$$

to complete the construction of the model; in general this is a hard task, the inverse of what we usually do in statistical mechanics.

If the expectation values that we measure are $\langle\sigma_\mathrm{i}\rangle$ and $\langle\sigma_\mathrm{i}\sigma_\mathrm{j}\rangle$, then the corresponding maximum entropy distribution can be written as

$$P(\vec{\sigma}) = \frac{1}{Z}\exp\left[\sum_{\mathrm{i}=1}^{M} h_\mathrm{i}\sigma_\mathrm{i} + \frac{1}{2}\sum_{\mathrm{i}\neq\mathrm{j}}^{N} J_\mathrm{ij}\sigma_\mathrm{i}\sigma_\mathrm{j}\right], \tag{599}$$

where the 'magnetic fields' $\{h_\mathrm{i}\}$ and the 'exchange couplings' $\{J_\mathrm{ij}\}$ have to be set to reproduce the measured values of $\{\langle\sigma_\mathrm{i}\rangle\}$ and $\{C_\mathrm{ij}\}$. This of course is an Ising model with pairwise interactions among the spins. What is crucial is that this model emerges here not through hypotheses about the network dynamics, but rather as the least structured model that is consistent with the measured expectation values. The mapping to the Ising model is a mathematical equivalence, not an analogy, and the details of the model are specified by the data.

The emergence of the Ising model is an attractive aspect of the maximum entropy construction. But, there is no obvious reason why real biological networks should have this maximum entropy property. Indeed, one might guess that there are complicated, higher order correlations which are important for the function of the network, and these will be missed by a maximum entropy model built only from pairwise correlations. It thus came as a surprise when it was found that these models really do provide an accurate description of the full correlation structure in the vertebrate retina as it responds to naturalistic stimuli. This has led to considerable interest in the use of these models more generally for the description of real neural networks; for details, see Appendix A.8.

Problem 124: Maximum entropy model for a simple neural network. Imagine that we record from N neurons and we find that all of them have the same mean rate of spiking, $\bar{r}$. Further, if we look at any pair of neurons, the probability of both spiking in the same small window of duration $\Delta\tau$ is $p_c = (\bar{r}\Delta\tau)^2(1+\epsilon)$. We want to describe this network as above, with Ising variables $\sigma_\mathrm{i} = +1$ for spiking and $\sigma_\mathrm{i} = -1$ for silence.

(a.) Show that

$$\langle\sigma_\mathrm{i}\rangle = -1 + \bar{r}\Delta\tau \tag{600}$$

$$C_\mathrm{ij} \equiv \langle\sigma_\mathrm{i}\sigma_\mathrm{j}\rangle - \langle\sigma_\mathrm{i}\rangle\langle\sigma_\mathrm{j}\rangle = 4\epsilon(\bar{r}\Delta\tau)^2. \tag{601}$$

(b.) Since all neurons and pairs are equivalent, the maximum entropy model consistent with pairwise correlations has the simpler form,

$$P(\vec{\sigma}) = \frac{1}{Z}\exp\left[h\sum_{\mathrm{i}=1}^{M}\sigma_\mathrm{i} + \frac{J}{2}\sum_{\mathrm{i}\neq\mathrm{j}}^{N}\sigma_\mathrm{i}\sigma_\mathrm{j}\right], \tag{602}$$

which is just the mean field ferromagnet (assuming that J is positive). If N is large, one might expect that there is a 'thermodynamic limit' in which quantities like energy and entropy become extensive, proportional to N. Show that this requires scaling of the coupling, $J = J_0/N$. With this scaling, derive the relationship between the derivatives of $\ln Z$ and the expectation values $\langle\sigma_\mathrm{i}\rangle$ and C_ij.

(c.) Some of you will be very familiar with the substitution tricks that we're about to use, others less so. To be sure, let me take you through the steps. We notice that the interactions are described by a term

$$\frac{J}{2}\sum_{\mathrm{i}\neq\mathrm{j}}^{N}\sigma_\mathrm{i}\sigma_\mathrm{j} = \frac{J}{2}\sum_{\mathrm{i,j}=1}^{N}\sigma_\mathrm{i}\sigma_\mathrm{j} - \frac{NJ}{2} = \frac{J}{2}\left(\sum_{\mathrm{i}=1}^{N}\sigma_\mathrm{i}\right)^2 - \frac{NJ}{2}. \tag{603}$$

Thus the partition function can be written as

$$Z = \sum_{\vec{\sigma}}\exp\left[h\sum_{\mathrm{i}=1}^{M}\sigma_\mathrm{i} + \frac{J}{2}\sum_{\mathrm{i}\neq\mathrm{j}}^{N}\sigma_\mathrm{i}\sigma_\mathrm{j}\right] \tag{604}$$

$$= e^{-NJ/2}\sum_{\vec{\sigma}}\exp\left[h\sum_{\mathrm{i}=1}^{M}\sigma_\mathrm{i}\right]\exp\left[\frac{J}{2}\left(\sum_{\mathrm{i}=1}^{N}\sigma_\mathrm{i}\right)^2\right]. \tag{605}$$

Then the key step is to realize that

$$\exp\left[\frac{A}{2}(x)^2\right] = \int\frac{d\phi}{\sqrt{2\pi A}}\exp\left[-\frac{\phi^2}{2A} + \phi x\right]. \tag{606}$$

Applied to Eq (605) this allows us to write

$$Z = e^{-NJ/2}\sum_{\vec{\sigma}}\exp\left[h\sum_{\mathrm{i}=1}^{M}\sigma_\mathrm{i}\right]\exp\left[\frac{J}{2}\left(\sum_{\mathrm{i}=1}^{N}\sigma_\mathrm{i}\right)^2\right].$$

$$= e^{-NJ/2}\sum_{\vec{\sigma}}\exp\left[h\sum_{\mathrm{i}=1}^{M}\sigma_\mathrm{i}\right]\int\frac{d\phi}{\sqrt{2\pi J}}\exp\left[-\frac{\phi^2}{2J} + \phi\sum_{\mathrm{i}=1}^{N}\sigma_\mathrm{i}\right] \tag{607}$$

$$= e^{-NJ/2}\int\frac{d\phi}{\sqrt{2\pi J}}\exp\left[-\frac{\phi^2}{2J}\right]\sum_{\vec{\sigma}}\exp\left[(h+\phi)\sum_{\mathrm{i}=1}^{N}\sigma_\mathrm{i}\right]. \tag{608}$$

Now we see that the spins have decoupled, and you should be able to do the sum over states, $\sum_{\vec{\sigma}}$, inside the integral. Show that, with the scaling from (b.),

$$Z = e^{-NJ/2}\int\frac{d\phi}{\sqrt{2\pi J}}\exp\left[-NF(\phi;h,J_0)\right], \tag{609}$$

where the effective free energy $F(\phi;h,J_0)$ has no explicit N dependence.

(d.) Use steepest descent to approximate Eq (609) at large N. Derive an expression for $\ln Z$ which captures both the leading behavior ($\ln Z \propto N$) and the first two corrections.

(e.) To finish the construction of the model, we have to adjust h and J to match the measured means and pairwise correlations, Eq's

(600) and (601). Using the scaling required for a thermodynamic limit, is there a prediction for the N dependence of the correlation strength ϵ? This should bother you—ϵ is a quantity that is *measured* from pairs of cells, and shouldn't really depend on the number of cells in the network. Suppose we measure ϵ among more and more pairs of cells, so we have to describe larger and larger networks. Is it possible to have ϵ small and constant as $N \to \infty$? What conditions need to be met in order for this to happen?

The Hopfield model provides a scheme for the stabilizing multiple, discrete patterns of activity. But there certainly are situations in which the brain must hold a memory of a continuous variable. This is even less generic than the case of discrete attractors. In order to have a memory of a continuous variable, there must be (at least) a whole line or curve in state space along which the system can stop; if we think it terms of an energy landscape, then there must be one big valley, and the bottom of this valley must be precisely flat along one direction. Implausible as all this sounds, the brain really does hold memories of continuous variables, and it does so even in simple situations.

When you turn your head, cells in the semicircular canals, buried in the same bone as the cochlea, sense the rotational motion; this is called our "vestibular" sense. This angular motion input passes through the brain and drives a motor output which counter–rotates the eyes. This happens automatically, and is called the vestibulo–ocular reflex. You can demonstrate it for yourself by shaking your head from side to side as you read this text. If you are holding the book at arm's length, then in order to read you have to have your fovea—the $\sim 1^\circ$ wide area of highest image quality—focused on the words as you read them. If you move your head from side to side, and don't move your eyes to compensate, the text will blur. In fact, you (hopefully) have no trouble reading and shaking your head at the same time, suggesting that your eyes are being moved to compensate with an accuracy of better than $\sim 1^\circ$. When you are reading, of course, there are visual cues to help guide you eye movements, but it turns out that even if you close your eyes or sit in a dark room, seeing nothing, your eyes still counter–rotate to compensate for your head motions.

Problem 125: Mechanics of the semicircular canals. Give a problem to develop the simple mechanical model of the canal, explaining how one gets velocity sensitivity over a reasonable bandwidth. Use real dimensions of the canal (e.g., in humans) to get numbers out.

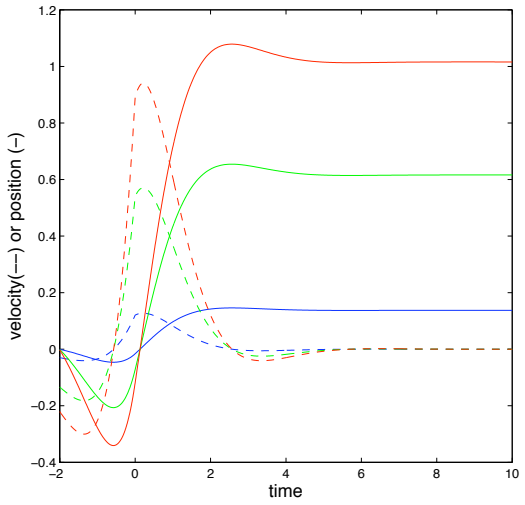

FIG. 125 Integration as memory for a continuous variable. Dashed lines show possible velocity signals, and the solid lines show corresponding position signals, obtained by integrating the velocity. After the transient inputs die away, the output of the integrator is stable for all time (a memory) and can take on any real value.

There is a subtlety of the vestibulo–ocular reflex, however. If we relax all the muscles to our eyes, then they rotate to a resting position in which we are looking more or less straight ahead (as defined by where our nose is pointing). Thus, if we turn out head to the right and stop, we need to keep tension on the eye muscles to be sure that they don't drift away from where we were looking before we turned. That is, to fully compensate for rotation of the head we need a signal related to the desired angular *displacement* of the eyes. But the vestibular system is an inertial sensor, driven by angular accelerations; the mechanics of canal turn this into a velocity signal over a wide range of frequencies, but the sensors really have zero response to constant displacements. Thus, the brain needs to take a input related (at best) to head velocity, and generate an output related to head displacement—it has to integrate, where here 'integrate' has the literal meaning from calculus, rather than being a qualitative statement about the gathering of multiple signals. Although we don't usually think about it this way, an integrator is a device which, once the input signals die away, has a perfect memory for a continuous variable, as schematized in Fig 125. Although these properties of the integral are obvious mathematically, it it less obvious how to build a network of neurons that implements this mathematics.

Before continuing, it should be noted that the movement of our eyes is not the perfect integral of our head velocity. On a longer time scale, roughly thirty seconds, our eyes do drift back to a resting position if there is no further stimulus. But this time scale is very long compared with the natural time scales of individual neurons, perhaps by a factor of as much as $\sim 10^3$. Could this gap

be closed by an emergent long time scale in the network, resulting from a line or curve of fixed points?

Suppose that the activity of each neuron is described by a coarse–grained continuous variable, such as the rate r at which it generates action potentials. If we inject a current I into the neuron directly, we find that the rate changes, along some curve $r(I)$. Each spike arriving at a synapse onto cell i effectively injects current into that cell, but this current is smoothed by some dynamics which we will summarize by a time scale τ, and the spikes from cell j are weighted by the strength of the synapse W_{ij}. This suggests a simple model,

$$\tau \frac{dI_i}{dt} + I_i = \sum_j W_{ij} r(I_j) + I_i^{ext}, \tag{610}$$

where I_i^{ext} represents currents injected from outside the network, including from sensory inputs. Typical examples of the response function $g(I)$ are sigmoids, threshold linear relations, and other monotonic functions. [add figure to show some examples of $g(I)$?]

What would it mean for the dynamics of Eq (610) to be an integrator? At the very least, the dynamics has to look like an integrator in its linear response to inputs, so let's see how this is possible. Assume that in the absence of inputs, there is some steady state at which $I_i = I_i^*$. Then if we linearize around this, writing $I_i = I_i^* + u_i$, we have

$$\tau \frac{du_i}{dt} + u_i = \sum_j W_{ij} r'(I_j^*) u_j + I_i^{ext}. \tag{611}$$

As always with linear problems, we want to change coordinates so that matrices become diagonal. If we denote quantities in this new coordinate system by tildes, then we will have

$$\tau \frac{d\tilde{u}_n}{dt} + \tilde{u}_n = \Lambda_n \tilde{u}_n + \tilde{I}_n^{ext}, \tag{612}$$

where the eigenvalues are defined as solutions to

$$\sum_j W_{ij} r'(I_j^*) \psi_j^{(n)} = \Lambda_n \psi_j^{(n)}. \tag{613}$$

If one of the $\Lambda_n \to 1$, then along this direction we have simply

$$\tau \frac{d\tilde{u}_n}{dt} = +\tilde{I}_n^{ext}, \tag{614}$$

$$\Rightarrow \tilde{u}_n(t) = \tilde{u}_n(0) + \frac{1}{\tau} \int_0^t dt' \, \tilde{I}_n^{ext}(t'), \tag{615}$$

so that $\tilde{u}_n$ is the time integral of its inputs. Thus, being an integrator means arranging the matrix of synaptic strength so that it (is appropriate units) has a unit eigenvalue, which means that (at least in this one mode) the signals which are being received from other cells in the network perfectly balance the decay processes within each cell. This of course is a critical point in the dynamics—if the eigenvalue is larger than one, the dynamics become unstable, if it is less than one it is stable but an imperfect integrator. Only at the critical point is true integration achieved. If we are within ϵ of the critical point, the system will hold a memory for $\sim \tau/\epsilon$, so if we really need to span three orders of magnitude (or even two), then the adjustment to the critical point must be quite precise.

The language of eigenvalues and critical points makes precise our initial intuition that there is something highly non–generic about memory for a continuous variable. Most valleys have a single lowest point, and balls keep rolling downhill until they find it. Only at the critical point is there one perfectly neutral direction in the valley, along which the ball feels no force.

Problem 126: Details of the line attractor. [go through Seung (1996) to look for good questions about the linear algebra of the model]

The fact that the position of our eyes is the integral of the velocity signals from our semicircular canals, and that there is (apparently) a continuum of stable points where our eyes can sit, means that something like this description in terms of line attractors must be true for the system as a whole. Indeed this is more general: the fact that we (and other animals) can stabilize a continuously variable set of postures means that the combined dynamics of our limbs, muscles, sensors and brain must have a line or manifold of attractors. It is more challenging to point to a particular part of the system—e.g., a particular sub–network of neurons in one part of the brain—and claim that the dynamics of this subsystem must have a line or attractors.

Seeing a model which explains things but only for particular choices of parameters makes us uneasy, as in our previous examples in this chapter. But in this case, we know that the relevant parameters—synaptic strengths—are adjustable, because this is how we learn. Also, we know that if we make errors, then under normal conditions (with the lights on) these errors are literally visible as slippage of the image on our retina as we turn out heads. There must be some way to use this error signal to adjust the synaptic weights and tune the network to its critical point. Does this happen?

To test the idea that the brain tunes the dynamics of the integrator circuit to its critical point, Major et al did a seemingly simple but beautiful experiment using goldfish, which also exhibit oculomotor integration. Essentially they built a planetarium for the goldfish, and then

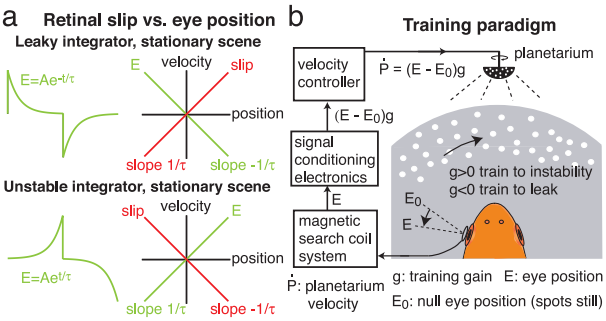

FIG. 126 Schematic of the "planetarium experiment," from Major et al (2004a). At left, the dynamics of a leaky (top) or unstable (bottom) integrator are evident as exponential decay or growth of eye position. This can be analyzed by plotting eye velocity vs eye position, revealing a straight line with a sign that indicates stability or instability, and a slope that measures the time constant of the system. At right, the planetarium setup, in which eye movements are monitored and fed back to movements of the surrounding scene.

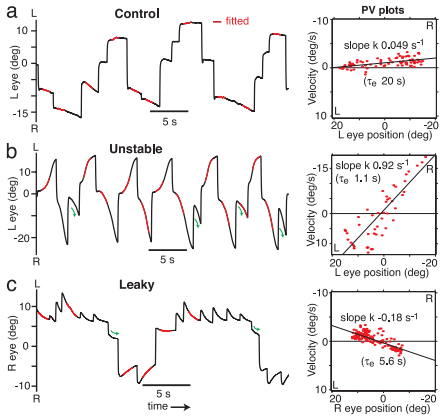

FIG. 127 Results of the planetarium experiment, from Major et al (2004a). At the top (a), control experiments showing the eye trajectories and position vs velocity plots before exposure to the feedback system in the planetarium. Note that the time constant of the system is $\sim$ 20 s. After exposure to feedback which should "teach" the system to be unstable (b) or leaky (c), trajectories and position vs velocity plots show the expected behaviors, with time constants for growth or decay on the oder of $1 - 5$ s.

coupled the rotation of this 'world' to their eye movements, as in Fig 126. Under normal conditions, when the eyes move by an angle θ, this is equivalent to the world moving the other way by the same angle. But if we give an additional rotation, we can create a situation in which the world slips on the retina even when the integrator network is set correctly. If the system in fact continuously uses slip signals to tune the system, this will drive a mistuning, either toward stability or instability. If we remove the feedback, we should then see that the fish can no longer stabilize its gaze, with the eyes either quickly relaxing to their resting position or exploding wildly away from rest, needing correction by frequent saccades.

The quick summary is that all of what we expect to see is observed experimentally, as summarized in Fig 127. Importantly, one can record from neurons in the relevant circuit and demonstrate that the detuning of the behavioral integration is mirrored by changes in the dynamics of persistent neural firing. While this does not prove that the line attractor scenario is correct, it does show that the long time scale of memory exhibited by the oculomotor integrator is the result of an active tuning process which uses visual feedback as a control signal. In this way, non–generic behavior of the system is learned, robustly.

Need an introduction.

Amit 1989: *Modeling Brain Function: The World of Attractor Neural Networks.* DJ Amit (Cambridge University Press, Cambridge, 1989).

Cooper 1973: A possible organization of animal memory and learning. LN Cooper, in *Collective Properties of Physical Systems: Proceedings of Nobel Symposium 24,* B Lundqvist & S Lundqvist, eds, pp 252–264 (Academic Press, New York, 1973).

Hertz et al 1991: *Introduction to the Theory of Neural Computation* J Hertz, A Krogh & RG Palmer (Addison Wesley, Redwood City, 1991).

Hopfield 1982: Neural networks and physical systems with emergent collective computational abilities. JJ Hopfield, *Proc Nat'l Acad Sci (USA)* **79,** 2554–8 (1982).

Hopfield 1984: Neurons with graded response have collective properties like those of two–state neurons. JJ Hopfield, *Proc Nat'l Acad Sci (USA)* **81,** 3088–3092 (1984).

McCulloch & Pitts 1943: A logical calculus of ideas immanent in nervous activity. WS McCulloch & W Pitts, *Bull Math Biophys* **5,** 115–133 (1943).

Now talk about connections to real neurons

Hebb 1949: *The Organization of Behavior: A Neuropsychological Theory.* DO Hebb (Wiley, New York , 1949).

James 1892: *Psychology: The Briefer Course.* W James (Henry Holt and Company, 1892). There is a modern edition from Dover Publications (New York, 2001), based on the 1961 abridged version from Harper and Row (New York, 1961).

Lorente de No 1938: Analysis of the activity of the chains of internuncial neurons. R Lorente de No, *J Neurophysiol* **1,** 207–244 (1938).

Introduce more powerful stat mech approaches

Amit et al 1985: Spin–glass models of neural networks. DJ Amit, H Gutfreund & H Sompolinsky, *Phys Rev A* **32,** 1007–1018 (1985).

Amit et al 1987: Statistical mechanics of neural networks near saturation. DJ Amit, H Gutfreund & H Sompolinsky, *Ann Phys* **173,** 30–67 (1987).

Crisanti et al 1986: Saturation level of the Hopfield model for neural networks. A Crisanti, DJ Amit & H Gutfreund, *Europhys Lett* **2,** 337–341 (1986).

Pointers to experiments on persistent activity (need more!).

Funahashi et al 1989: Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. S Funahashi, CJ Bruce & PS Goldman–Rakic, *J Neurophysiol* **61,** 331–349 (1989).

Fuster & Alexander 1971: Neuron activity related to short–term memory. JM Fuster & GE Alexander, *Science* **173,** 652–654 (1971).

Prut & Fetz 1999: Primate spinal interneurons show pre–movement instructed delay activity. Y Prut & EE Fetz, *Nature* **401,** 590–594 (1999).

The idea of using maximum entropy models to think about correlations in networks of neurons arose from a very practical problem—if we observe correlations among pairs of neurons, should we be surprised if we observe, for example, three or four neurons generating action potentials simultaneously? For continuous variables, we can separate different orders of correlations quite simply (recall the idea of cumulants in statistics, or "connected diagrams" in field theory). For discrete variables, pairwise correlations imply higher order correlations, even without any further assumptions. One touchstone for this idea is in statistical mechanics—recall that the usual Ising model has only interactions between two spins at a time, but when we coarse grain this model to give the Landau–Ginzburg Hamiltonian, we generate ϕ^4 interaction terms, so that the magnetization ϕ (which is a spatially smoothed version of the original spins) must have nontrivial fourth order correlations [should give some standard ref]. Schneidman et al (2003) showed how one could use the maximum entropy construction to generalize the idea of connected correlations to discrete variables. [Be careful here .. maybe push more into Appendix A.8?]

Schneidman et al 2003: Network information and connected correlations. E Schneidman, S Still, MJ Berry II & W Bialek, *Phys Rev Lett* **91,** 238701 (2003).

When we set out to use the maximum entropy method to analyze the responses of real neurons in the vertebrate retina, we expected we would "clean out" the pairwise correlations and uncover the higher order effects which were responsible for the known tendency of many neurons to fire simultaneously (Schnitzer & Meister 2003). The surprising result was that the pairwise Ising model provides a very accurate description of the combinatorial patterns of spiking and silence in ganglion cells of the salamander retina as they respond to natural and artificial movies, and in cortical cell cultures (Schneidman et al 2006). After the initial success in the salamander retina, similarly encouraging results were obtained in the primate retina, under very different stimulus conditions (Shlens et al 2006, 2009), in visual cortex (Ohiorhenuan & Victor 2007, Yu et al 2008), and in networks grown in vitro (Tang et al 2008). Most of these detailed comparisons of theory and experiment were done for groups of $N \sim 10$ neurons, small enough that the full distribution $P_{\rm expt}(\{\sigma_i\})$ could be sampled experimentally and used to assess the quality of the pairwise maximum entropy model. Attempts to push to larger networks are described by Tkačik et al (2006, 2009) [and pointer to Appendix; see how far we can go before finalizing text].

Ohiorhenuan & Victor 2007: IE Ohiorhenuan & JD Victor, Maximum entropy modeling of multi-neuron firing patterns in V1. Proceedings of 2007 Cosyne conference; http://cosyne.org.

Schneidman et al 2006: Weak pairwise correlations imply strongly correlated network states in a neural population. E Schneidman, MJ Berry II, R Segev & W Bialek, *Nature* **440,** 1007–1012 (2006).

Schnitzer & Meister 2003: Multineuronal firing patterns in the signal from eye to brain. MJ Schnitzer & M Meister, *Neuron* **37,** 499–511 (2003).

Shlens et al 2006: The structure of multi-neuron firing patterns in primate retina. J Shlens, GD Field, JL Gaulthier, MI Grivich, D Petrusca, A Sher, AM Litke & EJ Chichilnisky, *J Neurosci* **26,** 8254-66 (2006).

Shlens et al 2009: The structure of large–scale synchronized firing in primate retina. J Shlens, GD Field, JL Gaulthier, M Greschner, A Sher, AM Litke & EJ Chichilnisky, *J Neurosci* **29,** 5022–5031 (2009).

Tang et al 2008: A Tang, D Jackson, J Hobbs, W Chen, JL Smith, H PAtel, A Prieto, D Petruscam MI Grivich, A Sher, P Hottowy, W Dabrowski, AM Litke & JM Beggs, A maximum entropy model applied to spatial and temporal correlations from cortical networks *in vitro*. *J Neurosci* **28,** 505–518 (2008).

Tkačik et al 2006: Ising models for networks of real neurons. G Tkačik, E Schneidman, MJ Berry II & W Bialek, arXiv:q–bio.NC/0611072 (2006).

Tkačik et al 2009: Spin glass models for networks of real neurons. G Tkačik, E Schneidman, MJ Berry II & W Bialek, arXiv:0912.5409 [q–bio.NC] (2009).

Yu et al 2008: S Yu, D Huang, W Singer & D Nikolic, A small world of neuronal synchrony. *Cereb Cortex* **18,** 2891–2901 (2008).

[Refs for oculomotor integrator]

Major et al 2004a: Plasticity and tuning by visual feedback of the stability of a neural integrator. G Major, R Baker, E Aksay, B Mensh, HS Seung & DW Tank, *Proc Nat'l Acad Sci (USA)* **101,** 7739–7744 (2004).

Major et al 2004b: Plasticity and tuning of the time course of analog persistent firing in a neural integrator. G Major, R Baker, E Aksay, HS Seung & DW Tank, *Proc Nat'l Acad Sci (USA)* **101,** 7745–7750 (2004).

Robinson 1989: Integrating with neurons. DA Robinson, *Ann Rev Neurosci* **12,** 33–45 (1989).

Seung 1996: How the brain keeps the eyes still. HS Seung, *Proc Nat'l Acad Sci (USA)* **93,** 13339–13344 (1996).

E. Perspectives

The exploration of fine tuning vs. robustness in biological systems encourages us to think beyond models for this or that particular system. To ask whether some function requires fine tuning of parameters, we imagine that the system we are looking at is just one member in a class of possible systems. Whatever the answer to our initial questions, this effort at generalization clearly is an important step on the path to a physicist's view of life.

When we think about individual proteins, generalization is easy—proteins are polymers, and there is a natural class of molecules that can be built from the same

monomers, but with different sequences. When we think about a biochemical or genetic network, with many interacting protein molecules, it seems natural to generalize to a class of networks that has the same topology, but different parameters on each node or link. The ion channels in a single neuron provide an important example of a network of interacting proteins, where the interactions are mediated by the (global) transmembrane voltage and, importantly, experiments on single channel molecules serve to validate the equations describing what happens at each node. Finally, for networks of neurons, the fact that the strengths of synaptic connections are 'plastic' makes it natural to think about classes of networks that have the same topology of connections among neurons, but with different strengths. In all of these cases, we can see that the generalization to a class of networks is not just a useful theoretical construct, but also something which has meaning in the life or evolution of the organism.

In the extreme, "robustness" would mean that functional behavior is largely invariant over the whole class of networks. If this really is the case, then we should be able to choose networks at random and have them function. This is essentially the strategy employed by many groups searching for robustness in biochemical networks, and long before this there was a serious exploration of neural networks with randomly chosen strengths of synaptic connections among all the cells, using analytic methods borrowed from the dynamical theory of spin glasses. In the context of neural networks, the model with random connections indeed behaves chaotically, which seems odd, although it has been suggested that in the absence of other inputs this is the right answer—sensory inputs serve to drive the network out of the chaotic phase into an ordered state. For biochemical and genetic networks chaos seems less generic, but to obtain functional behavior without adjusting parameters there is general agreement that the topology of the network must be chosen carefully. There are several open questions here. Why is chaos not more common in large networks of biochemical reactions? What is the boundary between changing parameters (e.g., make the rate on one particular chemical reaction smaller) and changing topology (setting that rate exactly to zero)? To speak precisely about what will be typical of a randomly chosen network, we need a measure on the space of parameters; is there a natural choice of this measure?

In most of the systems we have studied, the randomly chosen parameters do not correspond to functional behavior. Random amino acid sequences don't fold into functional proteins, randomly chosen numbers of ion channels will not generate the correct rhythms of electrical activity, and while random neural networks may perform some functions, they certainly don't provide for stable storage and recall of memories. In each of these cases there are mechanisms for tuning or selecting the functional regions of parameter space. In single neurons, adjusting the numbers of copies of different channels is a form of physiological adaptation, connecting electrical activity, intracellular messengers, and the control of gene expression. In neural networks, the strengths of synapses are adjusted during learning, and for some key processes this learning happens all the time—as perhaps is necessary if the behavior the system is trying to stabilize is very far from typical in the space of possible networks. Finally, for amino acid and DNA sequences, the "adjustment" to functional behavior occurs on evolutionary time scales.[74] In this context, we can think of adaption, learning and evolution as different mechanisms for accomplishing the same task, albeit on different time scales.

As we will see in Section IV.D, there is a sophisticated mathematical theory of learning, combining ideas from mathematics, computer science and statistical physics. In particular, in different contexts, this theoretical approach places bounds on what can be learned, and how quickly. If we see adaptation, learning and evolution as different approaches to the same problem, should there be a comparable theoretical framework limiting the speed of evolution, or the effectiveness of adaptation? For evolution there is, in the long run, an obvious external definition of correct functional behavior (successful reproduction), and for learning there are often external signals (as in the case of the oculomotor integrator) that define the goal of the learning process; in adaptation, how do cells "know" the correct behavior that they are trying to stabilize? In the models for regulation of ion channel densities that we discussed in Section III.B, this is (weakly) programmed into the cell by the parameters that define a target calcium concentration; is there a more general definition of when cells are getting things right? Are there, as with learning, limits on how precisely one can get things right if the system needs to adjust quickly?

To return to the opening remarks in this Chapter, we wanted to distinguish between the usual physicist's mistrust of explanations that rest on fine tuning of parameters, and some specifically biological notions of robustness or evolvability. Part of the motivation for robustness as a biological principle is the intuition that living organisms simply can't adjust parameters accurately enough to guarantee reliable, reproducible functions. I think this intuition turns out to be wrong—cells can and do exercise precise control over the numbers of molecules that they make, so that the absolute concentrations of relevant molecules *can be* reproducible from cell to cell (or, in the

[74] It is worth emphasizing that, in the immune system, there is a kind of accelerated evolution within individual organisms, and this serves to select a nontrivial distribution of sequences for the antibody molecules. See the discussion in Section III.A.

discussion of Section III.C, embryo to embryo) with high precision. I emphasize "can be," because one clearly cannot conclude that all concentrations or molecule counts will be reproducible in this way. Indeed, the example of ion channels makes clear that, in the natural parameter space for the cell, there are many different ways of achieving essentially the same function, and so there is no reason for the cell to control the number of copies of any one particular molecule very precisely; what is important are the tight correlations among variations in different molecule counts, and these correlations are often expected and observed to be nonlinear, even defining non–convex regions of parameter space.

The fact that they can exert precise control over the concentrations, or combinations of concentrations, of certain molecules does not solve all of the organism's problems. Most fundamentally, life as a cold blooded organism[75] means having to function across a range of conditions where all chemical reaction rates vary, often by an order of magnitude or more, with no guarantee that the different rates in a given network will scale together; for an example of this problem one need look no further than the familiar circadian rhythms, which have long been known to be invariant to temperature changes. At the same time, diversity of environments is one of the driving forces for speciation, so that (for example) the fruit flies that live at different latitudes, and hence different temperature ranges, are genetically distinguishable. Natural history abounds with stories of animals that seek out very special environments in which to lay their eggs, casting doubt on any glib statement that embryonic development is robust against environmental perturbations. Still, simple laboratory experiments demonstrate that many aspects of life are nearly invariant over a wide range of temperatures, much wider than we might expect from simple models.

Locating life on the spectrum between precisely controlled (rather than finely tuned) dynamics and some more generic or robust behavior is an incredibly important question. It touches, as we have seen, phenomena ranging from the states of single cells to the nature of our memories. It connects to theoretical ideas that have the potential to reach deeply into statistical physics and dynamical systems. Still, at the risk of making clear the limits of my own understanding, I would say that we are still searching for the best formulations of these questions. We need more experimental guidance about what features of behavior are robust against which variations, and we need evidence that organisms actually face these variations in their natural environment. On the theoretical side, we need more anchor points like the random heteropolymer and the random neural network, where we have a complete analytic understanding of what is expected in the truly generic case, and we need a statistical mechanics of systems with random parameters that allows us to deal with the case where these parameters have nontrivial distributions. These are substantial challenges.

---

[75] Most of the biomass on our planet is cold blooded, so this is a very general problem.

The idea of choosing parameters at random in biochemical networks was explored by Barkai and Leibler (1999) and by von Dassow et al (2000), among others, using simulations; see Sections III.C. Much earlier, Sompolinsky et al (1988) had analyzed the dynamics of random neural networks, identifying a transition between a stationary phase and a chaotic phase at a critical value of the typical synaptic strength. For attempts to connect these random networks to the behavior of cortex, see van Vreeswijk & Sompolinksy (1996, 1998). More recently, Rajan et al (2010) have emphasized that input signals can drive random networks across the transition between order and chaos, providing a possible new view of the nature of variability in cortical responses (Abbott et al 2010).

Abbott et al 2010: Interactions between intrinsic and stimulus–evoked activity in recurrent neural networks. LF Abbott, K Rajan & H Sompolinsky, in *Neuronal Variability and Its Functional Significance,* M Ding & D Glanzman eds, in press (Oxford University Press, Oxford, 2010); arXiv:0912.3832 (2009).

Rajan et al 2009: Input–dependent suppression of chaos in recurrent neural networks. K Rajan, LF Abbott & Sompolinsky, *Phys Rev E* **82,** 011903 (2010).

Sompolinsky et al 1988: Chaos in random neural networks. H Sompolinsky, A Crisanti & HJ Sommers, *Phys Rev Lett* **61,** 259–262 (1988).

van Vreeswijk & Sompolinsky 1996: Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274,** 1724–1726 (1996).

van Vreeswijk & Sompolinsky 1998: Chaotic balanced state in a model of cortical circuits. *Neural Comp* **10,** 1321–1371 (1998).

There is work on why chemical dynamics tends to be "simple." Should give pointers here. Maybe also some refs on circadian clocks, and speciation of flies in different latitudes.