

Physical limits of cells and proteomes

Ken A. Dill^{a,1,2}, Kingshuk Ghosh^{b,2}, and Jeremy D. Schmitz^{c,2,3}

^aLaufer Center for Physical and Quantitative Biology and Departments of Physics and Chemistry, Stony Brook University, New York, NY 11794;

^bDepartment of Physics and Astronomy, University of Denver, Denver, CO 80209; and ^cDepartment of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2008.

Contributed by Ken A. Dill, September 2, 2011 (sent for review August 10, 2011)

What are the physical limits to cell behavior? Often, the physical limitations can be dominated by the proteome, the cell's complement of proteins. We combine known protein sizes, stabilities, and rates of folding and diffusion, with the known protein-length distributions $P(N)$ of proteomes (*Escherichia coli*, yeast, and worm), to formulate distributions and scaling relationships in order to address questions of cell physics. Why do mesophilic cells die around 50°C? How can the maximal growth-rate temperature (around 37°C) occur so close to the cell-death temperature? The model shows that the cell's death temperature coincides with a denaturation catastrophe of its proteome. The reason cells can function so well just a few degrees below their death temperature is because proteome denaturation is so cooperative. Why are cells so dense-packed with protein molecules (about 20% by volume)? Cells are packed at a density that maximizes biochemical reaction rates. At lower densities, proteins collide too rarely. At higher densities, proteins diffuse too slowly through the crowded cell. What limits cell sizes and growth rates? Cell growth is limited by rates of protein synthesis, by the folding rates of its slowest proteins, and—for large cells—by the rates of its protein diffusion. Useful insights into cell physics may be obtainable from scaling laws that encapsulate information from protein knowledge bases.

cell biophysics | protein dynamics | protein stability | diffusion and folding | proteome modeling

What physical limits are imposed upon cells by the equilibria and kinetics of their proteomes? For example, how stable are the proteins in the proteome? What limits the speed of cell growth and replication? What determines the densities of proteins in cells? A cell's mass is largely protein [50% of the nonaqueous component (1–3)]. So, some behaviors of cells are likely to be dominated by the physical properties of its proteome—the collection of its thousands of different types of proteins. We develop here some biophysical scaling relationships of proteomes, and we use those relationships to make estimates of the physical limits of cell behavior. Our scaling relationships come from combining current databases of the properties of proteins that have been measured in vitro, with $P(N)$, the length distributions of proteins that are known for several proteomes. Some of the hypotheses we explore are not new; what is previously undescribed is the use of modern databases to make quantitative estimates of physical limits. A key point, previously also made by Therumalai (4), is that many physical properties of proteins just depend on N , the number of amino acids in the protein. We estimate the folding stabilities for mesophiles and thermophiles, the folding rates, and the diffusion coefficients of whole proteomes, and we compare these various rates at the end. We first consider the folding stabilities of proteomes.

Proteomes Are Marginally Stable to Denaturation

For at least 116 monomeric, two-state and reversible folding proteins there are calorimetric measurements of the folding stability, $\Delta G = G_{\text{unfolded}} - G_{\text{folded}}$, enthalpy ΔH , entropy ΔS , and heat capacity, ΔC_p . Data are now available for both mesophilic and thermophilic proteins. Taken over the full set of proteins,

these thermal quantities depend, remarkably, mainly just on the number, N , of amino acids in the chain. The relationship is simply linear. For both enthalpy and entropy the correlation with chain length is found to be the best at two reference temperatures, $T_h = 373.5$ K and $T_s = 385$ K, respectively (5–7). These linear dependencies for mesophilic proteins (59 from the list of 116 proteins) are well-fit by the expressions (see Fig. 1) (7):

$$\begin{aligned}\Delta H(T = T_h, N) &= (4.0N + 143)\text{kJ/mol} \\ \Delta S(T = T_s, N) &= (13.27N + 448)\text{J/(mol}\cdot\text{K)} \\ \Delta C_p(N) &= (0.049N + 0.85)\text{kJ/(mol}\cdot\text{K)} \\ \Delta G(N, T) &= \Delta H(T_h) + \Delta C_p(T - T_h) - T\Delta S(T_s) \\ &\quad - T\Delta C_p \log T/T_s \\ &= [(4.0N + 143) + (0.049N + 0.85)(T - T_h) \\ &\quad - T(0.01327N + 0.448) \\ &\quad - T(0.049N + 0.85) \ln(T/T_s)]\text{kJ/mol}.\end{aligned}\quad [1]$$

A simple generalization of Eq. 1 also accounts for the effects of pH, salts, and denaturants on protein stability, in addition to temperature (6).

So far, extensive studies have shown no other strong dependence of the thermal properties of folding stability. Stability does not appear to depend on amounts of secondary structure or types of tertiary structure, or numbers of hydrophobic amino acids or hydrogen bonds, or counts of salt-bridging ion pairs, for example. This is remarkable because other important properties of proteins—such as their native structures and biochemical mechanisms—often do depend strongly on details. Predicting protein stabilities is not, in general, improved by using knowledge of the native structure. So, we can predict the thermal stabilities of whole proteomes just from knowledge of their chain-length distributions. The distribution of chain lengths over the proteomes of 22 fully sequenced organisms is obtainable from genomics and proteomics databases and is known to be well approximated by a gamma distribution (8):

$$P(N) = \frac{N^{\alpha-1} e^{-N/\theta}}{\Gamma(\alpha) \theta^\alpha}.\quad [2]$$

The two parameters, α and θ , of the protein chain-length distribution are obtained from the measured mean and variance observed for a given proteome, using the expressions

$$\langle N \rangle = \alpha\theta \quad \text{and} \quad \langle (\Delta N)^2 \rangle = \alpha\theta^2,\quad [3]$$

Author contributions: K.A.D., K.G., and J.D.S. designed research and wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: dill@laufercenter.org.

²K.A.D., K.G., and J.D.S. contributed equally to this work.

³Present address: Department of Physics, Kansas State University, Manhattan, KS 66506.

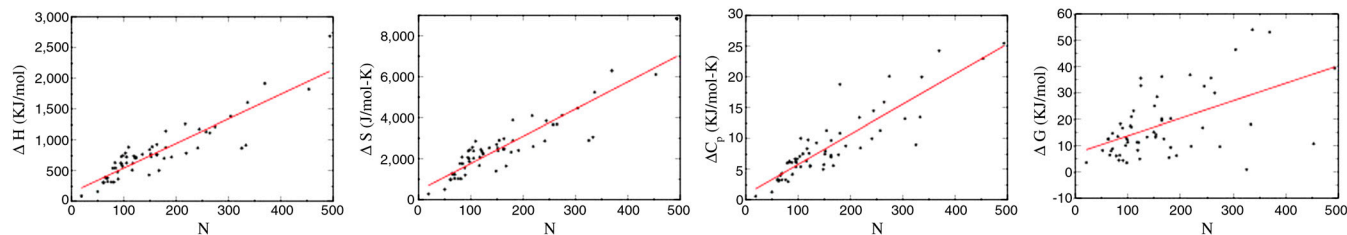


Fig. 1. Thermal protein properties depend linearly on chain length. (A) ΔH , enthalpy of unfolding at 100.5 °C, (B) ΔS , entropy of unfolding at 112 °C, and (C) ΔC_p , temperature independent heat capacity of unfolding. (D) Free energy of folding at 37 °C determined for 59 different proteins are plotted as a function of chain length, N . The red lines in A, B, and C are the linear regressions. The red line in D is the predicted ΔG ; it combines the best-fit values of changes in enthalpy, entropy, and specific heat on chain lengths obtained in A, B, and C. Reprinted from ref. 7, Copyright 2011, with permission from Elsevier.

where the brackets $\langle \dots \rangle$ indicate averaging over all the proteins in a proteome. For *Escherichia coli*, for example, the average protein length is $\langle N \rangle \approx 325$, and $\alpha = 2.33$ (8), so

$$P(N) = (8.5 \times 10^{-6})e^{-0.0072N}N^{1.33}. \quad [4]$$

Now, by combining Eq. 1 for the stability of a protein of length N with Eq. 2 for the distribution $P(N)$ of protein lengths in a proteome, we obtain an expression for the distribution of stabilities of a proteome's proteins, for a given temperature T , which is plotted in Fig. 2. Below, we use these expressions to explore how heat stresses affect cells.

Cells Are Sensitive to Small Changes in Temperature. It has been suggested that the reason that cells undergo a sharp heat-death transition is that their proteins all denature near the same temperature (9, 10), which is around 50 °C for *E. coli*; see Fig. 3. Fig. 2 gives a quantitative model and predicts that under normal physiological conditions the *E. coli* proteome is marginally stable. We find that a significant fraction of the proteome's proteins are poised near a denaturation catastrophe. Around 550 of *E. coli*'s 4,300 proteins are less than 3 kcal/mol stable to denaturation. This is a statement about the shape of the distribution function, not the average (11). The average protein at 37 °C is predicted to be stable by 6.8 kcal/mol. Rather, the marginal stability of the proteome comes from the broad left side of the distribution shown in Fig. 2; i.e., from the large numbers of proteins having small stabilities. This is a prediction (see also ref. 12); the actual shape of this distribution in vivo is not yet known from experiments. Note that although our conclusion is based on in vitro protein stability, experiments show that protein stabilities in vivo are not much different than in vitro (13, 14, 15).

Some cellular properties are highly sensitive to temperature. Upsifting by only 4 °C can often induce heat shock. This is puzzling because the thermal behaviors of materials are typically governed by RT , which means that a change of 4 °C would cause only a 1% effect. In contrast, the present model suggests that tempera-

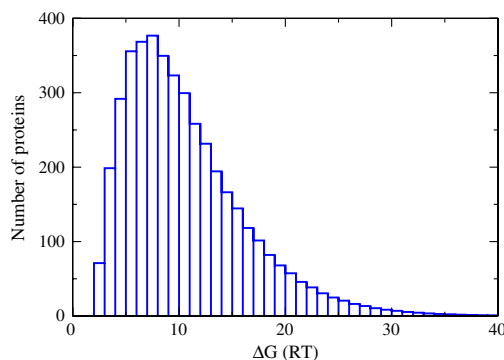


Fig. 2. Distribution of free energies of folding of the proteins in the proteomes of *E. coli* at 37 °C. The free-energy bin size is 1RT. Area under the curve equals the total number of proteins in the proteome (4,300 for *E. coli*).

ture effects are highly leveraged by the proximity of 37 °C to the proteome's denaturation phase boundary. The model shows that upshifting the temperature by only 4° from 37° to 41 °C destabilizes a proteome by nearly 16%. There are various biological properties of cells that show such high sensitivities, including the rates of evolution (16–18), some aspects of development (19, 20), and are the basis for some cancer treatments (21–23). Cells have evolved substantial machinery to handle the thermal denaturation of the proteome, including stress-responsive signaling pathways that transcriptionally up-regulate the proteostasis capacity of cells using chaperones, folding enzymes, and coupled disaggregation and degradation activities (24, 25). In addition, small changes in geological temperatures can cause ecological shifts in populations. We note that the stabilities reported here differ slightly from those given elsewhere (11) because of updated data (7).

Thermophilic Proteomes Are Different from Mesophilic Proteomes.

Above, we described the stabilities of *mesophiles*, organisms that grow at temperatures between 25 and 40 °C. The same approach can be applied to *thermophiles*, organisms that grow at temperatures higher than 40 °C. There is now a substantial database of two-state monomeric thermophilic proteins. What is the difference between the proteomes of mesophiles vs. thermophiles? A study of 59 mesophilic proteins and 57 thermophilic proteins indicates some systematic differences (7). Whereas the free energy of folding mesophilic proteomes is given by Eq. 1, we find that the free energy of folding thermophilic proteomes is given by

$$\begin{aligned} \Delta G(N,T)_{\text{thermo}} = & [(3.30N + 112) + (0.051N - 0.26)(T - 373.5) \\ & - T(0.0109N + 0.291) \\ & - T(0.051N - 0.26) \log\left(\frac{T}{385}\right)] \text{kJ/mol}. \quad [5] \end{aligned}$$

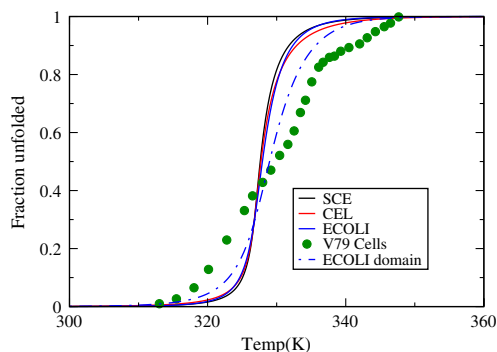


Fig. 3. The fraction of proteins that are unfolded in the proteomes of *E. coli*, yeast (SCE), and worm (CEL) as a function of temperature. Solid circles show experimentally measured fraction denatured proteins as a function of temperature for mammalian V79 cells using differential scanning calorimetry (9). The blue dashed line shows the results based on domain length distribution (69) in *E. coli* genome.

Comparing Eq. 1 with Eq. 5 and analyzing the database show that thermophilic proteomes have less positive entropies and enthalpies of unfolding, on average, than mesophilic proteomes (7). This suggests that the denatured states of thermophilic proteins may be more compact than the denatured states of mesophilic proteins. The reasons for these differences are not yet known, but may include differences in hydrophobic residues, disulfide bonds, electrostatic interactions (7, 26–36), different amino acid flexibilities (37–41), or loop deletions (42).

A Model for the Optimal Growth Temperatures of Cells. A cell has an optimal growth temperature. That is, cells grow fastest at one particular temperature, and slower at either higher or lower temperatures. The general view is that this optimum follows from a balance of two factors (7, 11, 43). On the one hand, biochemical processes have activation barriers, so heating a cell should accelerate its cellular metabolism. On the other hand, if a cell is heated too much, its proteome will denature. Here, we give a quantitative expression for cell growth vs. temperature, based on the stability expressions above. We model the growth rate, $r(T)$, of an organism as a function of temperature T using the expression

$$r(T) = r_0 e^{-\Delta H^\ddagger/RT} \prod_{i=1}^{\Gamma} f(N_i, T), \quad [6]$$

where r_0 is an intrinsic growth-rate parameter, ΔH^\ddagger is a dominant activation barrier, Γ is the number of proteins that are essential to growth, and

$$f(N_i, T) = \frac{1}{1 + \exp(-\Delta G(N_i, T)/RT)} \quad [7]$$

is the probability that the i th essential protein (having chain length N_i) is in the folded state. Following earlier work (43), Eq. 6 expresses the assumption that there are essential proteins, all of which, by definition, must be folded independently at a sufficient level for a cell to be viable. By definition, if any one essential protein is not folded, the cell is not viable. We compute $f(N_i, T)$ using Eq. 7 for the folding free energy $\Delta G(N_i, T)$. We simplify this expression by taking the logarithm and approximating the product by averaging over the entire proteome distribution (7, 11, 43). The result is

$$\langle \log r(T) \rangle = \log r_0 - \frac{\Delta H^\ddagger}{RT} + \Gamma \int \log f(N, T) P(N) dN, \quad [8]$$

where $P(N)$ is the proteome chain-length distribution and $f(N, T)$ is computed using $G(N, T)$ following Eq. 1 for mesophiles or Eq. 5 for thermophiles.

Fig. 4 shows that this model provides reasonable fits to growth curves for cells using two parameters, ΔH^\ddagger and Γ , for each cell type (7). It appears that cells have evolved to run as “hot” as possible for maximizing the speeds of their biochemical reactions, while remaining just cool enough to avoid a denaturation catastrophe. The reason this optimum growth temperature can be so close to the death temperature is because the denaturation of the proteome is so sharply cooperative (see Fig. 3).

Proteome stability models such as this may be useful for exploring other properties of cells. For example, Drummond et al. have proposed a relationship between the stability and rate of evolution of a protein. They argue that the most highly expressed proteins are optimized for stability; this enables those proteins to avoid misfolding and aggregation from sporadic mistranslation events. Stable proteins (which are highly expressed) have slower rates of evolution than less stable proteins (which have lower expression levels) (44–47).

Diffusional Transport of Proteins Within Cells

A limitation on the rates of a cell’s biochemical processes is the rate at which proteins can diffuse throughout the cell. For the purpose of computing diffusion rates, native proteins are often treated as being approximately spherical. The diffusion constant of a sphere is given by the Stokes–Einstein expression

$$D = \frac{RT}{6\pi\eta R_h}, \quad [9]$$

where η is the viscosity of the solvent in a dilute solution, RT is Boltzmann’s constant multiplied by temperature, and R_h is the radius of hydration of the sphere. Eq. 9 shows how the diffusion coefficient depends on the radius of the molecule. In the sections below, we simplify by expressing the radius, $R_h(N)$, of a native protein as a function of its number of amino acids, N , in the protein chain.

Native-Protein Sizes Scale as $R_n \approx N^{2/5}$. If native proteins were perfect spheres, their radii, R_n , should scale with chain length N as $R_n(N) \approx N^{1/3}$ (48–50). However, recent studies of the Protein Data Bank show instead a scaling approximately with $N^{2/5}$ (51):

$$R_n(N) = 2.24N^{0.392}, \quad [10]$$

based on an analysis of 37,000 protein structures, where N is the number of amino acids in the protein and R_n is given in angstroms. No major dependence was found of the native-state radius on the type or degree of secondary structure in the protein (51). The physical basis for the exponent of 2/5 is not fully clear, but can be rationalized either as being due to a balance between excluded volume and intermonomer elastic interactions (51) or because of the multidomain nature of many proteins. The difference in exponents between 1/3 and 2/5 would lead to negligible differences in fitting diffusion coefficients, of interest here; we use Eq. 10 below.

Now, in order to compute the diffusion properties, we need to obtain the hydrodynamic radius, R_h . We could adopt the expression for a perfect sphere, $R_h = (5/3)^{1/2} R_n \approx 1.3R_n$. However, in a survey of native proteins, Tyn and Gusek demonstrated that a better approximation is $R_h = 1.45R_n$ (52). Fig. 5 compares theory (combining this approximation with Eqs. 9 and Eq. 10) to experimental diffusion coefficients measured in vitro. The agreement is quite good. Combining the Tyn and Gusek expression with Eq. 10 and with the standard diffusion expression gives, for the time t required for a native protein of N amino acids to diffuse an rms distance of $\ell_0 = \langle x_0^2 \rangle^{1/2}$,

$$t = \frac{\ell_0^2}{6D} = \frac{c_0 \ell_0^2 \eta N^{0.392}}{RT}, \quad [11]$$

where $c_0 = \pi \times 1.45 \times 2.24 = 10.2 \text{ \AA}$. Combining Eq. 11 with the chain-length distribution, Eq. 4, gives the distribution of native-protein diffusion coefficients, at infinite dilution, for a proteome. However, to treat proteins in the cell, we must also account for the density of the molecular obstacles inside the cell. A protein will diffuse less distance in a given time in a crowded environment. We approximate the cell’s internal viscosity in terms of monodisperse hard spheres (53),

$$\eta = \eta_0 (1 - \phi/\phi_c)^{-2}, \quad [12]$$

where η_0 is the viscosity of water in the absence of crowding agents, and $\phi_c \approx 0.58$ is the critical concentration at which the glass transition occurs (54). The volume fraction of macromolecules in the cell is known to be $\phi \approx 0.2$ (55–57). Using Eq. 12 for the viscosity, Eq. 9 gives a reasonable approximation to diffusion rates for proteins in the cell for many purposes; see Fig. 6. We neglected attractive interactions between proteins. This is consis-

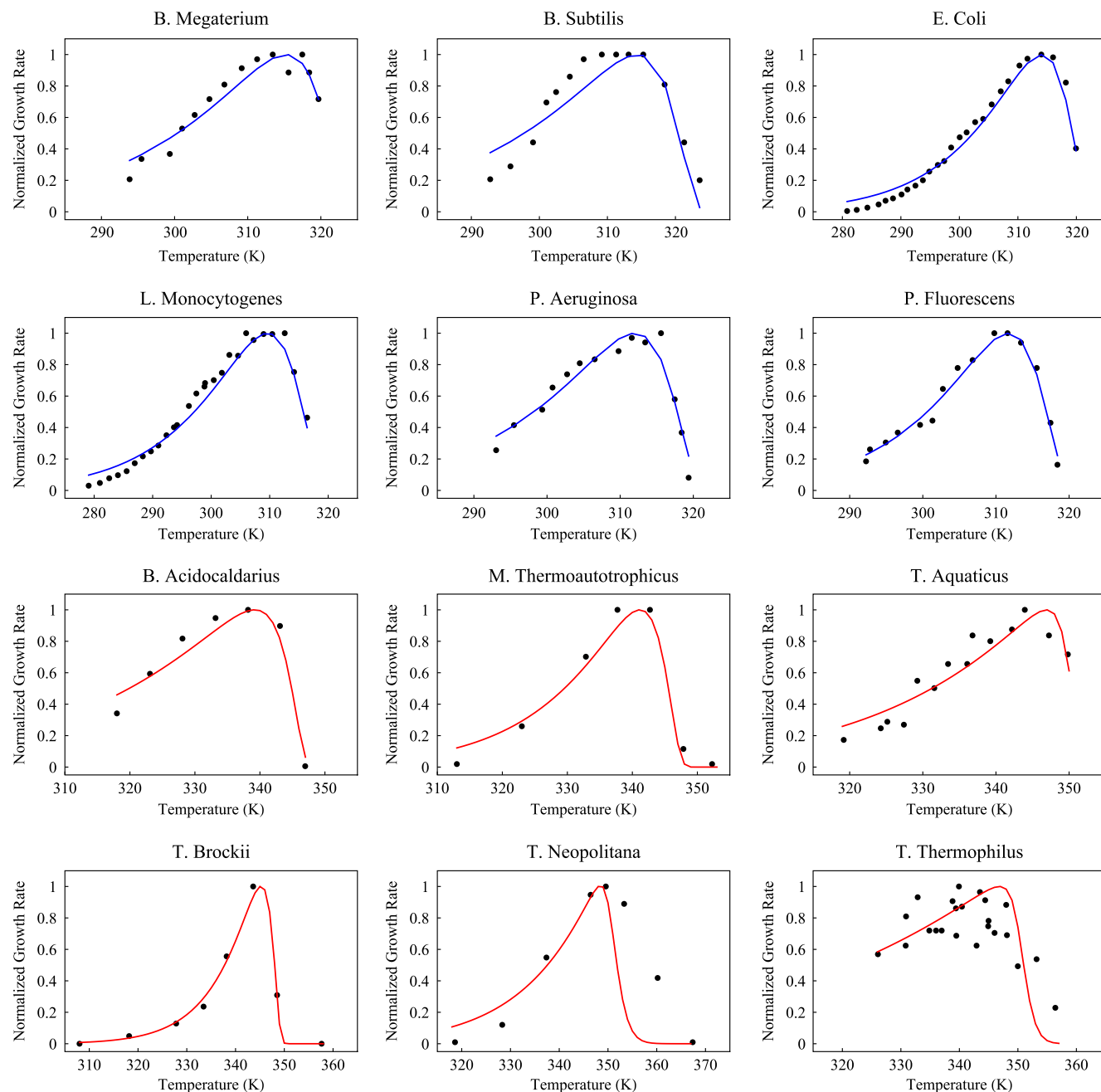


Fig. 4. Black circles denote growth rate as a function of temperature for species listed. On the Y axis, we plot growth rate normalized with respect to maximum growth rate, whereas on the X axis we plot temperature. Solid lines are fit to data from Eq. 8. Red graphs denote thermophilic species, whereas blue curves are for mesophiles. Sources of the data are given in table 2 in the supplemental material of ref. 7. Reprinted from ref. 7, Copyright 2011, with permission from Elsevier.

tent with the conclusions of Ando and Skolnick who argued for the dominance of excluded volume and hydrodynamic factors (58), but other work indicates that attractive forces can contribute more significantly (15).

Now, using Eq. 12, the expression for the diffusion time becomes

$$t = \frac{c_0 \ell_0^2 \eta_0 N^{0.39}}{RT(1 - \phi/\phi_c)^2}. \quad [13]$$

Using the proteome length distributions, $P(N)$, we compute the distribution of diffusion times for different cells, i.e., having different values of ℓ_0 (1 μm , 10 μm , and 100 μm). We now apply this model to a question of protein densities in cells.

Why Are Cells so Crowded with Proteins? We explore here the following hypothesis: Perhaps evolution maximizes the speed of biochemical reactions by varying the size of a cell so as to optimize the density of proteins inside the cell. On the one hand, if the cell's protein density were too low (cell size is too large), reactions would be slowed by the time required for the proteins to diffuse and collide with each other. On the other hand, if the cell's protein density were too high (cell size is too small), diffusion would be slow because of the crowded medium through which the proteins must move.

Here is a model of that hypothesis. Consider the rate at which diffusion-limited reactions occur. Such rates are proportional to the product of the concentration of the reactants and the diffusion constant. For example, a stationary reactant of radius a will

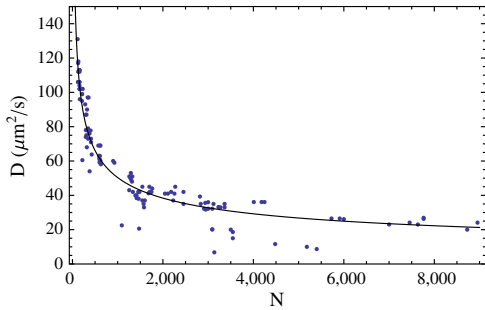


Fig. 5. Dependence of diffusion constant on chain length. Comparison of experimental in vitro diffusion constants (52) to Eq. 9 with the Tyn and Gusek approximation $R_D = 1.45R_n$. Molecular weights have been converted to amino acid numbers using $N = M_w/110$ Da.

react with mobile particles at a rate $r_d = 4\pi acD$, where c and D are the concentration and diffusion constant of the mobile reactants. Using the volume fraction ϕ instead of c , the rate is

$$r_d \sim cD \sim \phi \left(1 - \frac{\phi}{\phi_c}\right)^2. \quad [14]$$

Now, we treat this as an optimization problem. If we suppose that the number of protein molecules is fixed, then evolution can change a cell's protein density by inversely changing the cell's volume. A large cell would have a low concentration, and a small cell would have a high concentration. To find the optimal (maximum) reaction rate, we set the derivative to zero,

$$\frac{dr_d}{d\phi} = \left(1 - \frac{\phi}{\phi_c}\right)^2 - 2\left(1 - \frac{\phi}{\phi_c}\right)\left(\frac{\phi}{\phi_c}\right) = 0. \quad [15]$$

r_d has a maximum at $\phi = \phi_c/3$ (59). Because $\phi_c = 0.58$ for hard spheres (54), this model predicts the optimal density for maximal reaction rates would be $\phi = 0.19$. This is close to the protein densities ($\phi \approx 0.2$) observed in cells (55–57).

Protein Folding Kinetics

The physical dynamics within a cell is also limited by the rate at which its proteome folds. In a well-known observation, Plaxco, Simons, and Baker (PSB) (60) showed that the folding rates of two-state proteins correlate with the topology of the native structure: Proteins that have more local structure (such as helices) fold faster than proteins that have more nonlocal structures (such as β -sheets).

Alternatively, in equally good agreement with the experimental data, Thirumalai explained that folding rates correlate with chain length N , through the form $k_f \sim \exp(-N^{1/2})$ based on theory of random systems (61). Ouyang and Liang (62) looked at a broader set of proteins than PSB, including multistate folders.

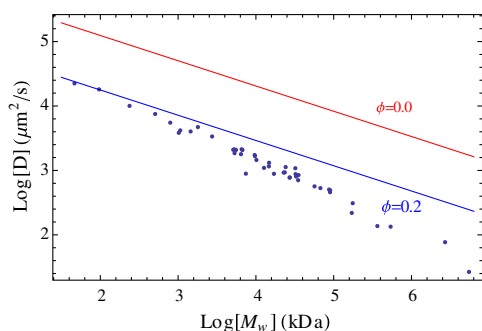


Fig. 6. Effect of crowding on diffusion constants. Computed diffusion constants at infinite dilution ($\phi = 0$) and in the presence of crowding compared to the steric interaction simulation of McGuffee and Elcock (15).

Other studies also give folding rates (63) under unifying conditions and study rates as a function of chain length (64–68). Here we use the Thirumalai expression to best fit the Liang data, giving

$$\ln k_f = 16.15 - 1.28N^{0.5}, \quad [16]$$

where k_f has units of second^{-1} . This shows that folding rates, $k_f(N)$, are relatively well predicted (correlation coefficient of 0.78) simply as a function of chain-length N (see Fig. 7). Hence, we can combine this expression with $P(N)$, the proteome chain-length distribution, to predict the folding rates of proteomes over whole proteomes (here, taken at the denaturation temperature for each protein). Fig. 8 shows the predicted folding rate distribution for *E. coli* using the domain distribution (11, 69). We use this treatment below for comparing the time scales of various processes in the cell.

Comparing the Time Scales of Dynamical Processes in the Cell. Now, we combine the various rate distributions above. Fig. 8 shows some rate distributions (for *E. coli*): protein folding, protein diffusion across the cell, rates of biochemical reactions (uncatalyzed and catalyzed), and rates of protein synthesis assuming a translation rate of 15 amino acids per second (70). The vertical bar on the right shows *E. coli*'s roughly 20-min replication time under fast-growth conditions (1). By comparing different rates, Fig. 8 gives some useful insights about the physical limits on cells.

First, it can be inferred that *E. coli* has evolved to replicate at speeds approaching its maximum possible “speed limit.” Here is the argument. Imagine two limiting cases. On the one hand, *E. coli* could have evolved so that each cell had only a single ribosome. On the other hand, *E. coli* could have evolved so that every cell is full of ribosomes. In the former case, *E. coli* would replicate very slowly; in the latter, *E. coli* would replicate very rapidly. Here are simple numerical estimates for these two conceptual limiting cases: (i) *Slowest possible, series replication.* If each *E. coli* cell contained only a single ribosome, that ribosome would have to copy every protein in the cell, one at a time, in series, before the cell could reproduce. Taking the single protein-replication time to be around 20 s (see Fig. 8 and ref. 2), it would take 2 y for each *E. coli* cell to replicate all of its 3 million protein molecules. (ii) *Fastest possible, parallel replication.* *E. coli* can replicate much faster than this because the cell has multiple ribosomes. A cell can “parallel process” its protein synthesis. Imagine the maximum conceivable parallelization: If every protein molecule in *E. coli* were a ribosomal protein, then each ribosome (a 55-protein complex of about 7,400 amino acids) would need only to synthesize its own 55 proteins in series, not 3 million. In this limit, *E. coli* could replicate in around 8 min [7,400/(15 aa/s)]. This is a simple “ballpark” estimate, not expected to be good to better than a factor of two or three. In reality, *E. coli* is able to replicate

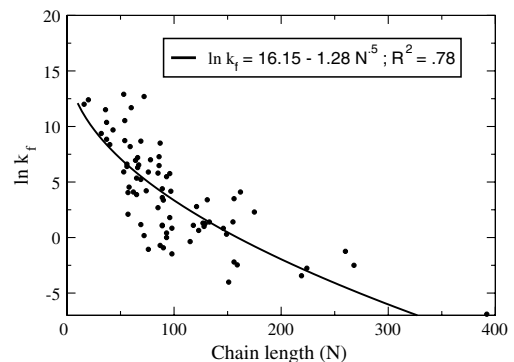


Fig. 7. Chain-length dependent fit (Eq. 16 using result from ref. 61) to the folding rates of 80 proteins including two-state and multistate folders (62). Data in circles and fit in solid line.

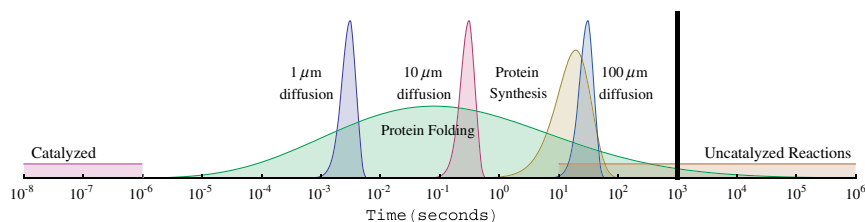


Fig. 8. Smoothed histograms (logarithmically spaced bins) of the time distributions for protein folding, translation, and diffusion compared to the range of reaction rates reported by ref. 72. Folding time distribution was calculated using domain length distribution in the proteome of *E. coli* and Eq. 16. The vertical line indicates 20 min, roughly the replication speed limit for *E. coli*.

at a speed that is exceedingly close to this maximum possible physical speed limit. This simple estimate also illustrates how a cell's physical limitations can contribute to its evolutionary limitations.

Second, enzymes are highly effective. Catalyzed reactions are rarely rate-limiting for the cell and are much faster than diffusion and folding. A key kinetic bottleneck in *E. coli* is the folding of the cell's slowest proteins, which happens on a time scale commensurate with protein synthesis.

Third, what limits the sizes to which cells can grow? One argument is that cell sizes are limited by surface-to-volume ratios. Cells that are too large in volume may be limited by the rate at which nutrients are taken up, which in turn is limited by the cell's membrane surface area. Another possible factor limiting cell growth is shown in Fig. 8: the rates at which a protein can diffuse across it. Eqs. 11 and 13 show that every factor of 10 increase in linear size of a cell leads to a factor of 100 increase in the time required for proteins to diffuse across the cell's length. Fig. 8 shows that for cells that are 100 μm in linear dimension, protein-diffusion transit times across a cell can become rate-limiting. So, although small cells are rate-limited by protein synthesis, large cells may also be limited by diffusional transport of proteins around the cell. Interestingly, prokaryotes are typically a few micrometers in size, eukaryotes are a few tens of micrometers, and there are very few cells much larger than 100 μm (2). Biology's need to move molecules around inside larger cells may have con-

tributed to the evolution of compartmentalization into organelles, active transport machinery, nonthermal transport (71) systems, and chaperones for speeding up protein folding.

Summary. We use simple scaling and distribution relationships, derived from recent databases, to describe some physical properties of proteins in cellular proteomes. The data show that many properties of proteins, including their sizes, stabilities, folding rates, and diffusion coefficients, depend simply on the chain length N . We combine data with the chain-length distributions $P(N)$ of proteins in cellular proteomes to make simple quantitative estimates of cell properties. We find that the death temperatures of cells, around 50 $^{\circ}\text{C}$, coincide with a denaturation catastrophe of the proteome. We find that the high protein densities in cells (around 20% volume fraction) can be rationalized on the basis that it maximizes biochemical reaction rates. And, we find that the diffusion and folding of the slowest proteins may impose limits to cell growth speeds and sizes.

ACKNOWLEDGMENTS. We thank Adrian Elcock for providing us with the simulation data for diffusion constants in Fig. 6. We thank Adrian Elcock, Dan Farrell, Jie Liang, Arijit Maitra, Wallace Marshall, Sefika Banu Ozkan, Alberto Perez, Larry Schweitzer, Dave Thirumalai, and Chris Voigt for helpful comments. We appreciate the support of National Institutes of Health Grant GM 34993.

1. Alberts B, et al. (2002) *Molecular Biology of the Cell* (Garland Science, New York).
2. Phillips R, Kondev J, Theriot J (2008) *Physical Biology of the Cell* (Garland Science, New York).
3. Nelson P (2008) *Biological Physics Energy, Information, Life* (Freeman and Company, New York).
4. Thirumalai D, O'Brien EP, Morrison G, Hyeon C (2010) Theoretical perspectives on protein folding. *Annu Rev Biophys* 39:159–183.
5. Robertson AD, Murphy KP (1997) Protein structure and the energetics of protein stability. *Chem Rev* 97:1251–1267.
6. Ghosh K, Dill KA (2009) Computing protein stabilities from their chain lengths. *Proc Natl Acad Sci USA* 106:10649–10654.
7. Sawle L, Ghosh K (2011) How do thermophilic proteins and proteomes withstand high temperature? *Biophys J* 101:217–227.
8. Zhang JZ (2000) Protein-length distributions for the three domains of life. *Trends Genet* 16:107–109.
9. Lepock JR (2005) Measurement of protein stability and protein denaturation in cells using differential scanning calorimetry. *Methods* 35:117–125.
10. Westra A, Dewey WC (1971) Variation in sensitivity to heat shock during cell-cycle of chinese hamster cells in-vitro. *Int J Radiat Biol* 19:467–477.
11. Ghosh K, Dill KA (2010) Cellular proteomes have broad distributions of protein stability. *Biophys J* 99:3996–4002.
12. Zeldovich K, Chen P, Shakhnovich EI (2007) Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proc Natl Acad Sci USA* 104:16152–16157.
13. Ghaemmaghami S, Oas TG (2001) Quantitative protein stability measurement in vivo. *Nat Struct Biol* 8:879–882.
14. Ignatova Z, Gierasch LM (2004) Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proc Natl Acad Sci USA* 101:523–528.
15. McGuffee SR, Elcock AH (2010) Diffusion, crowding protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput Biol* 6:e1000694.
16. Kennett JP, Stott LD (1991) Abrupt deep-sea warming, palaeoceanographic changes and benthic extinctions at the end of the Palaeocene. *Nature* 353:225–229.
17. Ivany LC, Patterson WP, Lohmann KC (2000) Cooler winters as a possible cause of mass extinctions at the Eocene/Oligocene boundary. *Nature* 407:887–890.
18. Allen AP, Gilooly JF, Savage M, Brown JH (2006) Kinetic effects of temperature on rates of genetic divergence and speciation. *Proc Natl Acad Sci USA* 103:9130–9135.
19. Ferguson MWJ, Joanel T (1982) Temperature of egg incubation determines sex in *Alligator mississippiensis*. *Nature* 296:850–853.
20. Webb GJW, Cooper-Preston H (1989) Effects of incubation temperature on crocodiles and the evolution of reptilian oviparity. *Am Zool* 29:953–971.
21. Weinberg RA (2007) *The Biology of Cancer* (Garland Science, New York).
22. Coffey DS, Getzenberg RH, DeWeese TL (2006) Hyperthermic biology and cancer therapies: A hypothesis for the "Lance Armstrong Effect". *JAMA* 296:445–448.
23. Vertrees RA, Zwischenberger JB, Boor PJ, Pencil SD (2000) Oncogenic *ras* results in increased cell kill due to defective thermoprotection in lung cancer cells. *Ann Thorac Surg* 69:1675–1680.
24. Balch WE, Morimoto RI, Dillin A, Kelly JW (2008) Adapting proteostasis for disease intervention. *Science* 319:916–919.
25. Dobson CM (2003) Protein folding and misfolding. *Nature* 426:884–890.
26. Cho JH, Sato S, Raleigh DP (2004) Thermodynamics and kinetics of non-native interactions in protein folding: A single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state. *J Mol Biol* 338:827–837.
27. Ge M, Xia, X-Y, Pan X-M (2008) Salt bridges in the hyperthermophilic protein Ssh10b are resilient to temperature increases. *J Biol Chem* 283:31690–31696.
28. Rosato V, Pucello N, Giuliano G (2002) Evidence for cysteine clustering in thermophilic proteomes. *Trends Genet* 18:278–281.
29. Clarke J, Hounslow AM, Bond CJ, Fersht AR, Daggett V (2000) The effects of disulfide bonds on the denatured state of barnase. *Protein Sci* 9:2394–2404.
30. Pace CN, Alston RW, Shaw KL (2000) Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Sci* 9:1395–1398.
31. Elcock AH (1998) The stability of salt bridges at high temperatures: Implications for hyperthermophilic proteins. *J Mol Biol* 284:489–502.
32. Kumar S, Tsai C, Nussinov R (2000) Factors enhancing protein thermostability. *Protein Eng* 13:179–191.
33. Kumar S, Ma B, Tsai C, Nussinov R (2000) Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. *Proteins* 38:368–383.
34. Dominy B, Minoux H, Brooks CL (2004) An electrostatic basis for the stability of thermophilic proteins. *Proteins* 57:128–141.
35. Zhou HX, Dong F (2003) Electrostatic contributions to the stability of a thermophilic cold shock protein. *Biophys J* 84:2216–2222.

