A simplified model of protein translocation

1 Biological problem

Mitochondria are double membrane organelle found inside most eukaryotic cells. In addition to supplying cellular energy, mitochondria are involved in other tasks, such as signaling, cellular differentiation, and cell death, as well as maintaining control of the cell cycle and cell growth.
Despite every mitochondrion has its own genome, the majority of processes that take place in these organelle require the presence of proteins that are synthesized in the cytosol, outside the mitochondrion (see Figure 4).
The import of such proteins inside the mitochondrion is made possible by the presence of several openings in its membrane, which are called pores.
As the protein needs to partially unfold in order to go through the pore, it is clear that the spontaneous import of one protein is not probable. For this reason cells have evolved a simple mechanism that allows the import to happen by mean of bounding it to a large protein (chaperone). The binding of this large protein is sufficient to help the translocation.
Why?

2 Solution

Qualitative idea

The only parameter we can use to describe the imported fragment of the protein (from here on also *the substrate*) is its length $n+1$ measured as the number of imported monomers. We define $F_0(n)$ as the free-energy of the imported chain not bounded to the chaperone protein, and $F_c(n)$ as the free-energy of the chain when bounded to the chaperone. As the contour length of the substrate increases of dl when a new residue has been imported, we can also define the importing force

$$f(n) = -\frac{F(n+1) - F(n)}{dl}. \tag{1}$$

In order for the translocation to be possible the importing force needs to be greater than the forces $g(n)$ (which we do not specify) that oppose the translocation.
As we know that the substrate does not spontaneously translocate, we deduce that the importing force before being bounded to the chaperone $f_0(n) = -\frac{F_0(n+1)-F_0(n)}{dl}$ is not sufficient to oppose $g(n)$, as instead is the case for the importing force after being bound to the chaperone as $f_c(n) = -\frac{F_c(n+1)-F_c(n)}{dl}$. Therefore our model should be at least able to predict that the difference

$$f_c(n) - f_0(n) = -\frac{[F_c(n+1) - F_0(n+1)] - [F_c(n) - F_0(n)]}{dl} \tag{2}$$

is positive.
In order to approach the problem we need to understand how the binding of a molecule can modify the free energy of the imported fragment. Without any molecular details about the system, we rely on three hypothesis in order to build our model. The first two assumption are the following:

1. **The chaperone protein is a sphere of radius R.** Without knowing the correct shape of the chaperone, this is the most trivial assumption

2. **The mitochondrion membrane is an ideal plane.** The typical size of mitochondria is $d_{mit} \approx 0.8 - 3.0\ \mu m$, several order of magnitude larger than the size of proteins. At the length scale of proteins the membrane can therefore be considered flat. for simplicity we choose the plane $z = 0$.

The first two assumptions are sufficient to understand how the free-energy of the system is modified upon binding. Indeed it is clear that all configurations of the chain in which the distance z between the end residue and the wall is less than R are not available to the substrate after the chaperone is bounded to it (see Figure 4). In short, the presence of a bounded protein reduces the conformational space of the substrate; the probability of observing a distance the end monomer of the substrate at a distance z from the membrane plane before the chaperone is bounded would remain the same after the bounding $iff\ z > R$, and will otherwise become zero.
Thus it is easy to compute

$$\int_R^\infty p(z,n) = \frac{Z_c}{Z_0} = \exp\left[-\beta\left(F_c(n) - F_0(n)\right)\right]. \tag{3}$$

Calculations

In order to estimate the force that the presence of a bounded protein exerts on the imported chain we therefore need to estimate $p(z,n)$. For this reason we make our last assumption:

3. **The imported protein can be approximated as a FJC, with bond length $dl = b$ and size $n+1$.** This assumption has been made for making the calculations possible. In general this would be a poor approximation, but notice that the only property of the chain that we need to be able to reproduce is the end-to-end distance distribution.

In the FJC model we then need to consider only those configurations which have the first residue very close to the membrane and whose monomers never cross the membrane plane (*i.e.* $z_i > 0\ \forall\ i$). The problem is easily solved by noticing the following :

1. **The distribution of the end-to-end distance of a FJC can be seen as the solution of a diffusion equation in which the diffusion coefficient $D = \frac{R^2_{max}}{6}$** (see Section 3.2)
2. **The requirement that the chain never touches the membrane is equivalent to imposing an absorbing boundary condition at $z = 0$ to the diffusion equation**
3. **The solution of the diffusion equation with one absorbing boundary can be obtained through the method of images** (see Section 3.3)

By applying the method of images we obtain:

$$p(z,n) \propto \sqrt{\frac{3}{2\pi n b^2}}\left\{\exp\left[-\frac{3(z-\epsilon)^2}{2nb^2}\right] - \exp\left[-\frac{3(z+\epsilon)^2}{2nb^2}\right]\right\}, \tag{4}$$

where ϵ is the distance between the membrane and the first residue of the chain, which is small but cannot be zero. Expanding the probability up to the first order in ϵ and normalizing:

$$p(z,n) = \frac{3}{nb^2}\ z\ \exp\left[-\frac{3z^2}{2nb^2}\right]. \tag{5}$$

In the presence of the chaperone we need to restrict the conformational space and compute

$$\int_R^\infty p(z,n) = \exp\left[-\frac{3R^2}{2nb^2}\right], \tag{6}$$

which can be used to obtain the free-energy change upon binding of a chaperone to a substrate of length $n+1$

$$F_c(n) - F_0(n) = k_B T\ \frac{3R^2}{2nb^2}. \tag{7}$$

Finally the increasing of the importing force due to the binding of a molecule is

$$f_c(n) - f_0(n) = k_B T \, \frac{3R^2}{2b^3} \left(\frac{1}{n} - \frac{1}{n+1} \right), \tag{8}$$

which is positive and decreasing with n (see Figure 4).
When bound in the proximity of the membrane (small n), the chaperone protein can pull the substrate against $g(n)$. But as the substrate continue to be imported, n increases and the force decreases until the import is stuck again. Another chaperone then needs to bind the substrate in the proximity of the pore to provide a the necessary force for the import to proceed.
The effect of the bounding of the chaperone to the system is purely entropic, as you can see from the fact that the force it provides is proportional to the temperature.

Final observations

The shape of the real chaperone protein which is involved in the import of protein through mitochondria membranes is seen in Figure 4. It appears evident that the chaperone is not approximated by an ideal sphere.
Despite this fact, we performed Molecular Dynamics (MD) simulations of the importing process by employing more realistic models for both the substrate and the chaperone. The increase of the importing force due to the binding is shown in Figure 4: a comparison with the prediction of the model here presented shows a good agreement between simulated and analytic results.

3 Appendix

3.1 A: Useful formulas

$$\int dx \; x \exp\left[-\frac{x^2}{2\sigma^2} \right] = -\sigma^2 \exp\left[-\frac{x^2}{2\sigma^2} \right] + const \tag{9}$$

3.2 B: Diffusion equation

Consider the process of creating, one monomer at a time, a FJC of final size maximum extension $R^2_{max} = Nb^2$, where $N+1$ is the number of total monomers of which the chain is composed and b the bond length. We introduce the variable $t_n = \frac{n-1}{N}$ and compute the probability of observing the monomer n at position $\boldsymbol{r}$. It is evident that $p(\boldsymbol{r}, t_n)$ is proportional to the probability that the previous monomer was at a distance b from $\boldsymbol{r}$. The proportionality factor is the inverse of the surface of a sphere of radius b, $4\pi b^2$. In formula:

$$p(\boldsymbol{r}, t_n) = \frac{1}{4\pi b^2} \int d\boldsymbol{r'} \; p(\boldsymbol{r'}, t_{n-1}) \delta(|\boldsymbol{r'} - \boldsymbol{r}|, b). \tag{10}$$

We introduce the orthonormal basis $\{\hat{x}_i\}_{i=1,2,3}$ and introduce the notation $\boldsymbol{r}_{x_i} = \boldsymbol{r} \cdot \hat{x}_i$ for describing the projection of the vector $\boldsymbol{r}$ onto the basis vectors $\hat{x}_i$.
If we take the limit $N \to \infty$, so that, at fixed R_{max}, $b \to 0$, it is possible to expand $p(\boldsymbol{r'}, t_{n-1})$ around $\boldsymbol{r}$:

$$\begin{aligned} p(\boldsymbol{r'}, t_{n-1}) =& p(\boldsymbol{r}, t_{n-1}) + (\boldsymbol{r'} - \boldsymbol{r}) \cdot \nabla p(\boldsymbol{r}, t_{n-1}) \\ &+ \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} (\boldsymbol{r'} - \boldsymbol{r})_{x_i} (\boldsymbol{r'} - \boldsymbol{r})_{x_j} \frac{\delta^2}{\delta x_i \delta x_j} p(\boldsymbol{r}, t_{n-1}) + \cdots . \end{aligned} \tag{11}$$

After substituting Eq. 11 into Eq. 10 and noticing that all the odd terms in $(\boldsymbol{r'} - \boldsymbol{r})_{x_i}$ of the expansion do not contribute to the integral, we obtain:

$$p(\boldsymbol{r}, t_n) - p(\boldsymbol{r}, t_{n-1}) = \frac{1}{2} \frac{b^2}{3} \; \nabla^2 p(\boldsymbol{r}, t_{n-1}). \tag{12}$$

The factor $\frac{b^2}{3}$ comes from the fact that $\sum_{i=1}^{3}(\boldsymbol{r'}-\boldsymbol{r})_{x_i}^2 = b^2$ by definitions, and therefore for isotropy reasons, each term in the sum must contribute for one third of b^2.
Finally we divide both the l.h.s. and the r.h.s. of the last equation by $t_n - t_{n-1} = \frac{1}{N}$ and, remembering we are in the limit of large N, we obtain

$$\frac{d}{dt}p(\boldsymbol{r},t) = D\,\nabla^2 p(\boldsymbol{r},t), \tag{13}$$

which is a diffusion equation with diffusion coefficient $D = \frac{R_{max}^2}{6}$.
The solution $p(\boldsymbol{r},t)$ in $\mathbb{R}^3$ with initial condition $p(\boldsymbol{r},0) = \delta(\boldsymbol{r}-\boldsymbol{r}_0)$ is

$$p(\boldsymbol{r},t_n) = \frac{1}{\left[4\pi\ Dt_n\right]^{\frac{3}{2}}}\exp\left[-\frac{(\boldsymbol{r}-\boldsymbol{r}_0)^2}{4Dt_n}\right], \tag{14}$$

and by substituting back $n-1 = Nt_n$ we obtain the usual result in polymer physics:

$$p(\boldsymbol{r},n) = \frac{1}{\left[\frac{2\pi}{3}\ b^2(n-1)\right]^{\frac{3}{2}}}\exp\left[-\frac{3(\boldsymbol{r}-\boldsymbol{r}_0)^2}{2b^2(n-1)}\right], \tag{15}$$

3.3 C: Method of images

The method of images (or method of mirror images) is a mathematical tool for solving differential equations, in which the domain of the sought function is extended by the addition of its mirror image with respect to a symmetry hyperplane. As a result, certain boundary conditions are satisfied automatically by the presence of a mirror image, greatly facilitating the solution of the original problem. You have already see this method in electrostatics, when you calculated the distribution of the electric field of a charge in the vicinity of a conducting surface.
Here wee will find a solution of the diffusion equation with initial condition $\theta(\boldsymbol{r},0) = \delta(\boldsymbol{r}_0)$ and boundary condition $\theta(0,t) = 0$.
For this purpose, consider the solutions $\phi(\boldsymbol{r},t;\boldsymbol{r}_0)$ and $\phi(\boldsymbol{r},t;-\boldsymbol{r}_0)$ of the diffusion equation without boundary conditions and initial conditions $\phi(\boldsymbol{r},0) = \delta(\boldsymbol{r}_0)$ and $\phi(\boldsymbol{r},0) = \delta(-\boldsymbol{r}_0)$ respectively. Without loss of generality, we choose a reference system in which $\boldsymbol{r}_0 = (0,0,r_0)$.
Because of symmetry, on the plane $z = 0$ these two solutions have the same value for every time t: $\phi(0,t;\boldsymbol{r}_0) = \phi(0,t;-\boldsymbol{r}_0)$. Therefore the difference $\theta(\boldsymbol{r},t) = \phi(\boldsymbol{r},t;\boldsymbol{r}_0) - \phi(\boldsymbol{r},t;-\boldsymbol{r}_0)$ satisfies the required boundary condition $\theta(0,t) = 0$.
It is trivial to show that any linear combination of solutions of the diffusion equation is itself a solution. Because of the uniqueness theorem (Cauchy's theorem), $\theta(\boldsymbol{r},t)$ is THE solution of our problem.

4 Images

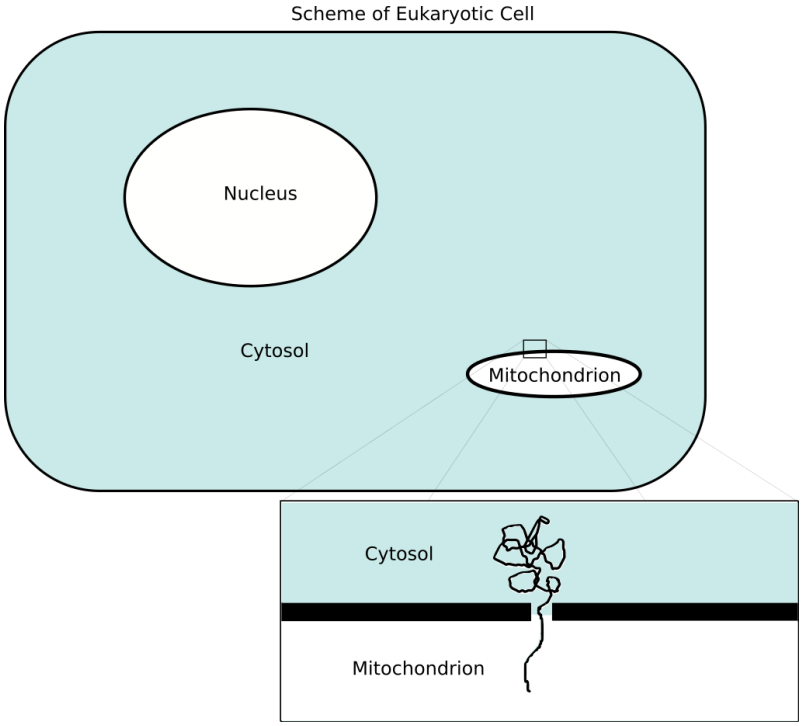

Figure 1: Highly simplified scheme of an Eukaryotic cell, which shows the nucleus (where DNA is stored), the cytosol (where protein are synthesized by rybosomes) and a mitochondrion. In the enlarged portion of the Figure a protein is translocating from the cytosol to the mitochondrion through a pore.

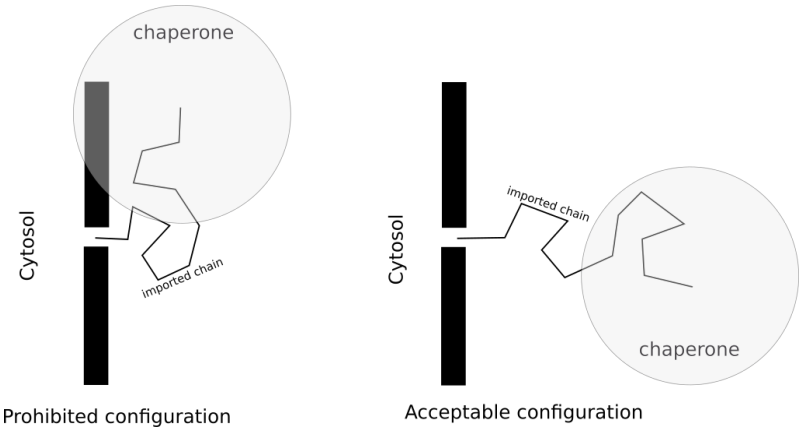

Figure 2: Some configurations of the imported portion of the chain may become prohibited because of the steric clash of the chaperone with the membrane.

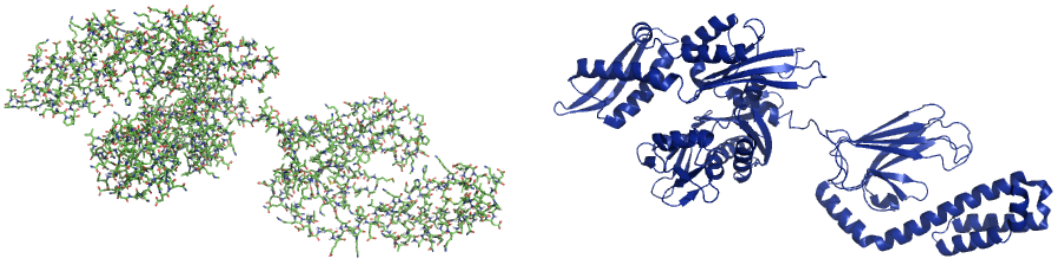

Figure 3: The chaperone involved in the import of protein in mitochondria: Hsp70 (pdb code: 2kho). Far from being a sphere, it is composed by two domains: a nucleotide binding domain (NBD) and a substrate binding domain (SBD). In the first representation (green) all atoms of the protein are visible. In the second one a cartoon is used to make the secondary structure visible. In both figures SBD is the rightmost domain. Dimensions of SBD: $\sim 60\text{Å} \times 30\text{Å} \times 20\text{Å}$. Dimensions of NBD: $\sim 55\text{Å} \times 50\text{Å} \times 35\text{Å}$.

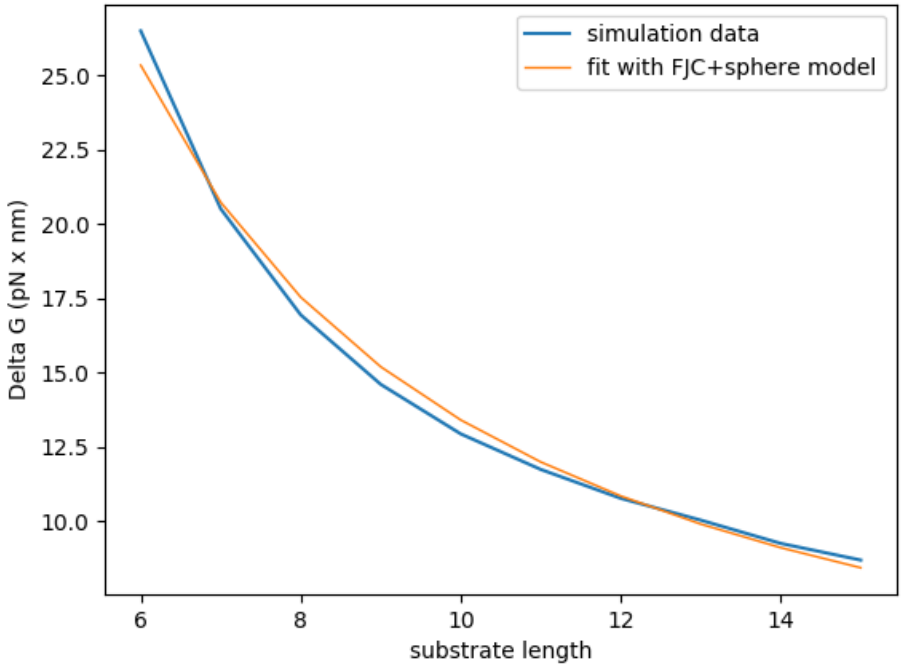

Figure 4: The graph shows $F_c(n) - F_0(n)$ as a function of n as obtained from Molecular Dynamics (blue) simulations. In orange the best rmsd fit obtained by using our model. The only parameter of the fit is the radius of the effective sphere, which results to be $R \sim 16.5\text{Å}$.