bioRxiv preprint doi: https://doi.org/10.1101/235689; this version posted December 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Statistical mechanics of cell decision-making: the cell migration force distribution

Haralampos Hatzikirou
Braunschweig Integrated Centre of Systems Biology and
Helmholtz Center for Infection Research,
Inhoffenstr 7, 38124 Braunschweig, Germany
haralampos.hatzikirou@helmholtz-hzi.de

***Abstract*—Cell decision-making is the cellular process of responding to microenvironmental cues. This can be regarded as the regulation of cell's intrinsic variables to extrinsic stimuli. Currently, little is known about the principles dictating cell decision-making. Regarding cells as Bayesian decision-makers under energetic constraints, I postulate the principle of least microenvironmental uncertainty (LEUP). This is translated into a free-energy principle and I develop a statistical mechanics theory for cell decision-making. I exhibit the potential of LEUP in the case of cell migration. In particular, I calculate the dependence of cell locomotion force on the steady state distribution of adhesion receptors. Finally, the associated migration velocity allows for the reproduction of the cell anomalous diffusion, as observed in cell culture experiments.**

I. Introduction

Cell decisions are responses of the cell's intrinsic mechanisms to extrinsic signals [1]. Such mechanisms are typically termed as signal transduction pathways. Typically, signal transduction occurs when an extracellular signalling molecule binds to a certain receptor creating a complex. In turn, this complex of molecules triggers a biochemical chain of events inside the cell [2]. These biochemical events can influence the epigenetic or even the genetic state of the cell. Then the new genomic state of the cell will be reflected in cell's transcriptome and translated into proteins that generate a phenotypic response. In general, cells encode extrinsic information into new intrinsic states and in turn decode the latter into phenotypic responses.

Concerning cell decision-making there are two fundamental issues: (i) typically there is a great deal of uncertainty in the intracellular dynamics involved in cell phenotypic decisions, and (ii) although there is a lot of work focused on single cell cell decision-making [] in the context of multicellular systems little is known. In particular, there are two schools of thought to address the single cell decision-making.

The first school thought is a *mechanism-driven* approach which follows the idea of epigenetic landscape reconstruction [3]. This dates back to Waddigton epigenetic landscape where cell phenotypes trickle down along a bifurcating landscape of "valleys" and "hills" (see Fig. 1). Reconstructing such landscapes requires very good knowledge of the underlying genetic, epigenetic or transcriptional dynamics and provides elegant mathematical formulations that allow for temporal predictions. Obviously, this approach suffers from the issue (i), i.e. we do not know all possible intracellular players and dynamics involved in phenotypic decision.

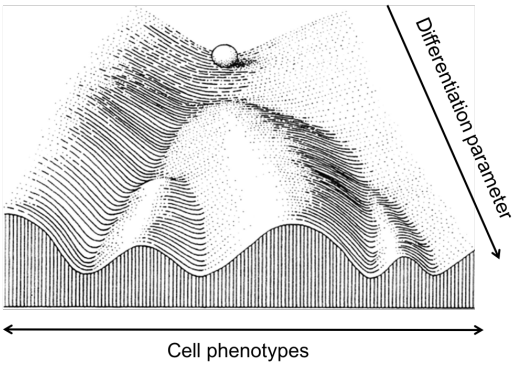

Fig. 1. Sketch of communication between cell and its microenvironment

The second is a *data-driven* approach involving ideas from statistical inference [4]. In principle, cell decisions are treated in a mechanism-blind way, as an input-output system, without requiring the knowledge of the intracellular components involved in phenotypic regulation.The intracellular dynamics are treated as a black box. The main goal is to optimize quantities such as mutual information and infer the corresponding phenotypic responses. The main problem with this approach is that ignores dynamics, i.e. does not allow for temporal predictions or analysis of the cells phenotypic dynamics.

Given the above problems, my goal is to identify a variational principle for cell decision-making that fulfils the following: (i) keeps the elegance and predictability of a dynamic description, and (ii) circumvent the details of intracellular dynamics as statistical inference. Here I propose the least microenvironmental uncertainty principle (LEUP) which I develop further in a statistical mechanics theory for cell decision making. Finally, I apply the LEUP theory into the cell migration problem and I calculate the corresponding force distribution and migration velocity.

II. The Bayesian cell decision-maker

Cells typically collect information from their microenvironment and encode it into tangible actions, i.e. phenotypic decisions. The latter involves the rearrangement of their intrinsic

bioRxiv preprint doi: https://doi.org/10.1101/235689; this version posted December 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

variables $\mathbf{x} \in \mathbf{X} \subset \mathbb{R}^n$, representing genes, RNA molecules, translational proteins, metabolites, membrane receptors, etc. Microenvironmental or extrinsic variables are denoted by $\mathbf{y} \in \mathbf{Y} \subset \mathbb{R}^n$ being ligand, chemicals, nutrients, cellular density or stress fields. In order to formalise this encoding process, we consider a Bayesian cell decision making:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \tag{1}$$

In particular, the extrinsic variable distribution $p(\mathbf{y}|\mathbf{x})$ represents the likelihood of mircoenvironmental information being collected by the cell. In turn, it is combined with the prior probability distribution function (pdf) of intrinsic variables $p(\mathbf{x})$ resulting into the posterior intrinsic variable distribution $p(\mathbf{x}|\mathbf{y})$. The latter represents the decision of the cell over the available information.

The collection of microenvironmental data, i.e. assesing the likelihood $p(\mathbf{y}|\mathbf{x})$, is energetically expensive. For instance, as shown in Fig. 2, sampling the cellular microenvironment involves the recruitment/degradation of different receptors (intrinsic state) and the corresponding binding to diffusible ligands (extrinsic state). As any physical system, cells would try to minimize energetic costs unless otherwise required. Therefore, I postulate that an energy-efficient cell would require the construction of the least informed/optimal prior[1]. This idea is rather straight forward in the case of differentiated cells, since heart cells, for instance, sufficiently know their heart microenvironment and they would fail living in an other one, e.g. skin.

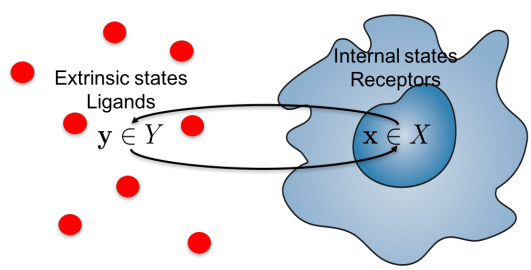

Fig. 2. Sketch of communication between cell and its microenvironment

III. A FREE-ENERGY PRINCIPLE: THE LEAST MICROENVIRONMENTAL UNCERTAINTY PRIOR

The above arguments could be translated in a statistical mechanics formulation. My central idea is that the cell decides on its phenotype by virtue of minimally sampling its microenvironment. As stated above, my goal is to optimize the cell's prior and at the same time minimize the total uncertainty of the system cell-microenvironment. This can be mathematically translated into finding the appropriate intrinsic state pdf $p(\mathbf{x})$ that minimizes the joint entropy of cell intrinsic and extrinsic variables

$$\min_{p(\mathbf{x})} S(\mathbf{X}, \mathbf{Y}), \tag{2}$$

where $S(\mathbf{X}, \mathbf{Y}) := -\int d\mathbf{x}d\mathbf{y}\, p(\mathbf{x}, \mathbf{y}) \ln p(\mathbf{x}, \mathbf{y}) = S(\mathbf{X}) + S(\mathbf{Y}|\mathbf{X})$.

The corresponding variational principle that follows reads:

$$\frac{\delta}{\delta p(\mathbf{X})}\Big[S(\mathbf{X}, \mathbf{Y}) - \lambda\Big(\int d\mathbf{x} p(\mathbf{x}) - 1\Big)\Big] = 0, \tag{3}$$

where the $\frac{\delta}{\delta p(\mathbf{X})}$ denotes the functional derivative operator and λ a Lagrange multiplier for the normalization constraint. The solution of eq. (3) is the equilibrium pdf:

$$p(\mathbf{x}) = \frac{e^{-\beta S(\mathbf{Y}|\mathbf{X}=\mathbf{x})}}{Z}, \tag{4}$$

where $Z = \int d\mathbf{x} e^{-\beta S(\mathbf{Y}|\mathbf{X}=\mathbf{x})}$ the corresponding normalization factor. The last equation is very interesting since it shows the equilibrium distribution of the cell's intrinsic states is the one that confers the least uncertainty of microenvironmental conditions. Please note that I introduced the parameter β as an "inverse temperature", that quantifies the compliance of a cell to LEUP. Please note that additional biological knowledge will be translated into Lagrange constraints of the free-energy minimisation.

To illustrate the above result, let us assume that the distribution $p(\mathbf{y}|\mathbf{x})$ follows the normal distribution $\mathcal{N}(\mu_y(\mathbf{x}), \sigma_y^2(\mathbf{x}))$. Then the mutual entropy reads

$$S(\mathbf{Y}|\mathbf{X}) = \ln(2\pi\sigma_y^2(\mathbf{x}))^{1/2},$$

and the eq. (4) respectively becomes:

$$p(\mathbf{x}) = \frac{\sigma_y^{-\beta}(\mathbf{x})}{\int d\mathbf{x}\, \sigma_y^{-\beta}(\mathbf{x})}. \tag{5}$$

This last equation is exactly the normalized square root of Fisher's Information for the second moment, i.e. $\sqrt{\mathcal{I}(\sigma_y(\mathbf{x}))} \propto 1/\sigma_y(\mathbf{x})$ [5]. Given the latter our equilibrium distribution (5) coincides with a normalized Jeffrey's prior distribution, for $\beta = 1$. Jeffrey's prior represents the definition of non-informative Bayesian prior in classical statistics [6].

Substituting the equilibrium pdf (4) into the cell intrinsic state entropy, we obtain:

$$S(\mathbf{X}) = -\int d\mathbf{x}\, p(\mathbf{x}) \ln \frac{e^{-\beta S(\mathbf{Y}|\mathbf{X}=\mathbf{x})}}{Z} = -\beta S(\mathbf{Y}|\mathbf{X}) + \ln Z. \tag{6}$$

The above is a fundamental thermodynamic identity of LEUP where the first term denotes the system's total phenotypic internal energy $U = \langle S(\mathbf{Y}|\mathbf{X} = \mathbf{x})\rangle_{p(\mathbf{x})}$ which is the average microenvironmental uncertainty sensed by a cell. The last term represent a phenotypic free energy $F = \beta^{-1} \ln Z$.

[1] An informed prior is the one that requires the most sampling of new data. Here we are interested to least collection of data, i.e. the least informed prior.

bioRxiv preprint doi: https://doi.org/10.1101/235689; this version posted December 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

IV. Cell migration force probability distribution against integrin receptors

To illustrate the implications of the LEUP in a single cell-microenvironment system, we assume the problem of single cell migration (see Fig. 3). Cell membrane-expressed integrin receptors bind to the corresponding extra-cellular matrix (ECM) ligands. The formation of stable focal adhesion points (FA), i.e. bond clusters, is required to generate locomotion force. Cell migration starts when the force at the front of the cell is larger than the force at the rear, inducing the rupture of the rear cell bonds. Integrin receptors play the role of cell's intrinsic state and the a density of ECM ligands the corresponding microenvironmental/extrinsic variable [7]. Here, I try to calculate the locomotion force distribution via the surface-expressing receptor distribution.

I employ the simplest FA dynamics model of receptor-ligand concentrations denoted by $x(t)$ and $y(t)$, respectively, and $c(t)$ corresponding to bond density :

$$\frac{dc}{dt} = k_+ y(x-c) - k_- c, \tag{7}$$

where k_+ the association and k_- the disassociation rate. Then, I calculate the steady state value of bond concentration:

$$c_{eq} = \frac{xy}{K_D + y}, \tag{8}$$

where the parameter $K_D = k_-/k_+$. It is well-known that the disassociation depends on the adhesion force F as $k_- = e^{F/c}$ and in the steady state one can prove (for details see in Schwartz et al [7]) that:

$$F_{eq} = plog(k_+/e)x, \tag{9}$$

where the function $plog$ is the solution of the equation $xe^x = a$. Therefore, the equilibrium force pdf is proportional to the steady state distribution of the receptors.

At this point, in order to calculate the equilibrium probability of receptors $p(x)$, I need to estimate the bond fluctuations of the system. The fluctuations estimated at equilibrium read:

$$\delta c_{eq} = \frac{dc_{eq}}{dy}\delta y = \frac{xK_D}{(K_D+y)^2}\delta y. \tag{10}$$

Using the Fourier transform $\delta\hat{c}_{eq}(\omega) = \int_0^T dt e^{-i\omega t}\delta c_{eq}(t)$, we can state that the standard deviation of c_{eq} is related to the power spectral density $|\delta\hat{c}_{eq}(\omega)|^2$ as

$$\langle \delta c_{eq}^2 \rangle = \frac{1}{2\pi}\int_{\mathbb{R}} |\delta\hat{c}_{eq}(\omega)|^2 d\omega.$$

The next goal is to calculate the steady state variation of complex formation. In this respect, we use the master equation as described in [8]:

$$\frac{dP_c}{dt} = k_+ y(x-(c-1))P_{c-1} + k_-(c+1)P_{c+1} \tag{11}$$

$$- [k_+ y(x-(c-1)) + k_-(c+1)]P_c, \tag{12}$$

where $P_c(t)$ the probability distribution of complexes and $c = 1, 2, ..., x-1$. The resulting probability generating function reads at the steady state:

$$G_{eq}(s) = \left(\frac{s+K_D y}{1+K_D y}\right)^x. \tag{13}$$

Then the complex distribution coincides with the Binomial one with parameter $(1+K_D y)^{-1}$. The variance is simply given by the following formula:

$$\langle \delta c_{eq}^2 \rangle = \frac{K_D xy}{(K_D+y)^2}. \tag{14}$$

Combining equations (10) and (14), we can obtain the steady state fluctuations of the ligand density:

$$\sqrt{\langle \delta y(x)^2 \rangle} = (K_D + y)\sqrt{\frac{y}{K_D x}}. \tag{15}$$

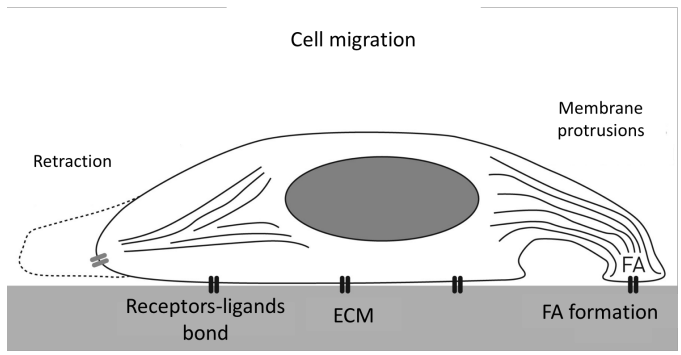

Fig. 3. Simplified description of cell migration processes.

Finally, assuming that ligands for a given receptor concentration follow a normal distribution and a "inverse temperature" β, the equilibrium pdf of receptor is provided by eq. (5):

$$p(x) = \frac{\langle \delta y(x)^2 \rangle^{-\beta/2}}{\int dx \langle \delta y(x)^2 \rangle^{-\beta/2}} \propto x^{\beta/2}, \tag{16}$$

which results into power law distributions for different values of β. Assuming normality of the ligand y distribution and c_{eq} (the binomial distribution converges to the normal one for large numbers of receptors), then from the eq. (8) and for $K_D \gg 1$, we have $x \propto c_{eq}/y$. One can easily prove that the ratio of two Gaussian distributions results into fat-tailed Cauchy distribution which scales the data, i.e. $1/x^2$. LEUP would fit the corresponding data for $\beta = -4$.

The above result states that the cell locomotion force pdf is proportional to the power law distribution of the adhesion receptors

$$p(F_{eq}) \propto x^{\beta/2}.$$

In particular it implies that there is a finite probability for cells to exert high locomotion forces. In principle, the presence of power laws would support the long-standing hypothesis that biology is poised at criticality [9].

bioRxiv preprint doi: https://doi.org/10.1101/235689; this version posted December 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

It is very interesting to see the impact on the cell locomotion velocity, where using the overdamped Newton's law reads:

$$v(t) \propto F_{eq} \propto x^{\beta/2}. \tag{17}$$

Now assuming that $x \propto t^a$, where $a \geq 0$, the velocity becomes

$$v(t) \propto t^{a\beta/2}.$$

It is know that for constant velocity $a\beta = 0$ the cell would be a random walker moving diffusively. However for $a\beta \neq 0$, one may recover the anomalous diffusion as observed in cell motility experiments [10] (see Fig. 4).

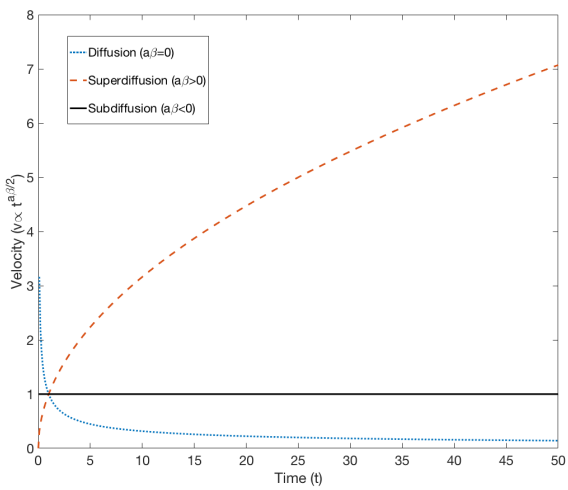

Fig. 4. Velocity temporal scaling $a\beta$ defines the different diffusion regimes.

As stated above, in the low association limit $K_D \gg 1$ the inverse temperature should be $\beta = -4$. Then the cell migration velocity decays in time as

$$v(t) \propto t^{-2a},$$

for $a \geq 0$. This is rather intuitive since for $K_D \gg 1$ very few bonds would be created. As a consequence, the cell would move sub-diffusively until it gets stopped.

V. Discussion

In this article, I present a novel statistical theory for cell decision-making, the so-called least microenvironmental uncertainty principle (LEUP). This is translated into a variational principle that minimized the joint entropy of cell intrinsic states and the corresponding microenvironental states. In turn, I develop the statistical mechanics theory for equilibrium cell decision-making, where I derive thermodynamic-like relationships. In turn, I apply the theory in a simplified model of cell migration and I calculate the locomotion force distribution. Finally, the resulting force distribution recovers the observed single cell anomalous diffusion.

In the present paper, I develop only the equilibrium theory. However, the minimization of the LEUP free energy allows us to formulate cell decision-making dynamics. This can be in the form of Monte-Carlo simulations or Langevin-like stochastic differential equations. In a future work, I will elaborate on this direction.

One could note that this is not the first effort to use information theoretic arguments to understand cell fate determination. Similar approaches have been undertaken by other works, such as [11], [12], [13], [14]. However, in another paper I discuss the connection of all these approaches with LEUP [15].

Finally, LEUP is particular interest since it allows for the inference of cellular intrinsic states/phenotypes by means of the local microenvironental entropy or fluctuations. This allows for computation of the cellular states without the explicit understanding the underlying mechanisms. The sole knowledge of the extrinsic variable distribution is enough. Therefore, we could apply the LEUP to problems where phenotypic regulation is unclear or unknown. One direct application would be on the regulation of migration/proliferation plasticity or Go-or-Grow in the context of glioblastoma tumors [16], [17].

References

[1] I. Golding, "Decision making in living cells: lessons from a simple system." *Annual review of biophysics*, vol. 40, pp. 63–80, 2011.

[2] J. E. Purvis and G. Lahav, "Encoding and decoding cellular information through signaling dynamics," *Cell*, vol. 152, no. 5, pp. 945–956, 2013.

[3] H. H. Chang, M. Hemberg, M. Barahona, D. E. Ingber, and S. Huang, "Transcriptome-wide noise controls lineage choice in mammalian progenitor cells," *Nature*, vol. 453, no. 7194, pp. 544–547, 2008.

[4] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, "Information transduction capacity of noisy biochemical signaling networks." *Science (New York, N.Y.)*, vol. 334, no. 6054, pp. 354–8, 2011.

[5] B. R. Frieden and P. M. Binder, *Physics from Fisher Information: A Unification*, vol. 68, 2000.

[6] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms David J.C. MacKay*, vol. 100, 2005.

[7] U. S. Schwarz and S. A. Safran, "Physics of adherent cells," *Reviews of Modern Physics*, vol. 85, no. 3, pp. 1327–1381, 2013.

[8] D. a. Lauffenburger and J. J. Linderman, *Receptors: models for binding, trafficking, and signaling*, vol. 39, no. 1, 1996.

[9] T. Mora and W. Bialek, "Are Biological Systems Poised at Criticality?" *Journal of Statistical Physics*, vol. 144, no. 2, pp. 268–302, 2011.

[10] P. Dieterich, R. Klages, R. Preuss, and A. Schwab, "Anomalous dynamics of cell migration," *Sciences-New York*, vol. 105, no. 2, pp. 459–463, 2008.

[11] R. C. Smith and B. D. , "Information-Theoretic Approaches to Understanding Stem Cell Variability," *Current Stem Cell Reports*, 2017.

[12] B. D. Macarthur and I. R. Lemischka, "Statistical mechanics of pluripotency," *Cell*, vol. 154, no. 3, pp. 484–489, 2013.

[13] B. R. Frieden and R. A. Gatenby, "Information dynamics in living systems: Prokaryotes, eukaryotes, and cancer," *PLoS ONE*, vol. 6, no. 7, 2011.

[14] W. Bialek, *Biophysics : searching for principles*. Princeton University Press, 2012.

[15] H. Hatzikirou, "Cell decision-making theory of multicellular systems," *in preparation*, 2018.

[16] K. Böttger, H. Hatzikirou, A. Voss-Böhme, E. A. Cavalcanti-Adam, M. A. Herrero, and A. Deutsch, "An Emerging Allee Effect Is Critical for Tumor Initiation and Persistence," *PLOS Computational Biology*, vol. 11, no. 9, p. e1004366, 2015.

[17] H. Hatzikirou, D. Basanta, M. Simon, K. Schaller, and A. Deutsch, "'Go or Grow': the key to the emergence of invasion in tumour progression?", *Mathematical medicine and biology*, pp. 1–17, 2010.