

From developmental metaphor to quantitative framework: A review of Waddington landscapes and their applications

Journal Title
XX(X):1–4
©The Author(s) 2016
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/

SAGE

John Vastola¹ and William R. Holmes²

Abstract

Achieving precise control of stem cell differentiation remains one of the most challenging open problems in biology. Some experimental reasons for this will be discussed elsewhere in this thematic issue; however, an important theoretical reason precise control remains elusive is that it is not clear how best to think about one cell changing into another type of cell. The epigenetic landscape view of differentiation, where one imagines a cell as a ball rolling around a rugged pasture full of hills and valleys, offers one way to think; however, it has proven difficult to make this conceptual picture mathematically precise, partly because there are many inequivalent proposals for accomplishing this. In this review, we discuss how to think about the landscape, its different mathematical formulations, the implications of those formulations, and applications to various areas of biology. We conclude with the practical issues currently preventing the landscape from being a useful, predictive tool for understanding differentiation.

Keywords

Class file, $\LaTeX 2_{\epsilon}$, SAGE Publications

Introduction

Changing some skin cells from a patient into heart cells or brain cells—at least, cheaply and in high enough volumes to be clinically useful—has proven to be harder than understanding fundamental properties of nature (like the anomalous magnetic moment of the electron) to ten decimal places, or using astronomical observations to estimate how long ago the universe began. *Why?* Isn't our encyclopedic knowledge of molecular biology, medicine, mathematics, and computation enough?

There are many, many, many good reasons, some of which will be discussed elsewhere in this thematic issue on iPSC differentiation: difficulties with experimental set-up, culturing, automation, data collection and analysis, and so on. But there are still other reasons, which prevent even the best-designed equipment in the world from being enough to solve the problem on its own. They boil down to the following: how should we *think* about iPSC differentiation? More generally, how should we think about one cell changing into another type of cell?

One popular approach is to take the so-called *epigenetic landscape* (or *Waddington landscape*) view of cell identity. Usually, the differences between one type of cell and another type of cell in the same organism are not genetic, but *epigenetic*: different proteins are expressed more or less highly, different transcriptional regions tend to be silenced or activated (PICK SOME BIOLOGICAL EXAMPLES HERE)...In this view, a well-equipped and sufficiently patient researcher could in principle change any cell into any *other* type of cell, provided that they adjusted the expression levels of each protein, made sure all important methylations were present, and so on.

The standard way of imagining one cell changing into another type of cell, in this view, is to picture a ball rolling around a rugged collection of hills and valleys. A ball in a valley corresponds to an epigenetic state which is somewhat hard to escape; if you move the ball slightly by making small epigenetic changes, the ball will roll back down into the valley. In other words, balls in valleys represent (possibly mature, possibly not) cell types or subtypes. A ball on a steep hill will roll down quickly until it reaches a valley; its epigenetic state is constantly changing.

This mental picture has some important corollaries. First, it suggests that differentiation is reversible, and that any cell can be *directly* reprogrammed into any other type of cell with sufficient effort.

Unfortunately, though the landscape has been helpful and influential as a metaphor and way of thinking about differentiation, it has been somewhat difficult to realize as a precise mathematical object. Part of the problem is that there are many inequivalent proposals for mathematically defining the landscape. Some of these proposals are equivalent *in certain limits*, as discussed by Zhou and Li [CITATION], but they are not equivalent in general. *Why?* How hard could it be to define a landscape?

¹Sunrise Setting Ltd, UK

²SAGE Publications Ltd, UK

Corresponding author:

Alistair Smith, Sunrise Setting Ltd Brixham Laboratory, Freshwater Quarry, Brixham, Devon, TQ5 8BA, UK.

Email: alistair.smith@sunrise-setting.co.uk

This issue is just one of the challenges that researchers face before the landscape can become a practical and predictive tool for understanding differentiation.

In this review, we will offer an explanation for why there are different landscapes—and why this should be expected, since the kind of landscape one is interested in will change depending on the question one is asking. We will also explain how to think about landscapes, how the various proposed landscapes are similar and different, and how they can be applied to understanding real biological systems. Finally, we conclude with some discussion of where landscape research might go next, given the important theoretical issues that remain to be addressed.

How to think about the landscape

For most biologists—especially experimental biologists and physicians—it is more helpful to know how to *think* about the landscape than it is to know how to perform technical landscape-related calculations. The intent of this section is to explore how to think about the landscape, partly by addressing common points of confusion.

In our experience, different researchers think about the landscape differently. But some ways are more appropriate than others, for reasons we will explain.

The standard conceptual picture

Picture a three-dimensional landscape with hills, valleys, and no other distinguishing features. In the standard metaphor, one imagines dropping the ball somewhere on this landscape, with the ball's starting position corresponding somehow to its current epigenetic state (what exactly this means will be discussed in more detail later). Next, the ball will move, mostly influenced by the contours of the landscape (moving quickly down a steep hill, slowly up a steep hill, and medium speed along level ground), but also influenced by intrinsic gene expression noise. In this metaphor, one might imagine intrinsic noise as wind that blows in random directions, and with random intensity.

Eventually, the ball will settle down into a valley. Because valleys are epigenetic states that are hard to escape, we identify them with cell types or subtypes. Because it is harder to (for example) change the epigenetic state of a mature cardiomyocyte than an iPSC, more mature cell types correspond to deeper valleys, while less mature cell types correspond to shallower valleys. Moreover, if two cell types are epigenetically 'close' to each other (we will discuss what this means later), then their corresponding landscape valleys are also close to each other.

Once the ball falls into a valley—that is, a cell becomes a specific cell type—it does not necessarily remain there forever, especially if that valley is relatively shallow. Because intrinsic noise causes the ball to move around somewhat randomly, there is a chance that it randomly receives enough of a push to be knocked out of the valley. From there, the ball may roll back into that valley, or perhaps roll into a completely different valley.

A ball can also be knocked out of a valley when an experimenter applies some external perturbation (by changing the cell's epigenetic state, or using a drug). This

can either be interpreted as the ball moving, or the landscape being reshaped so that the ball is more *likely* to move; which interpretation is more appropriate will be discussed in an upcoming section.

Problems with the standard conceptual picture

State-dependence of noise in general:¹

1. Increasing noise doesn't necessarily make jumping between attractors more likely.

Increasing noise makes jumping between attractors more likely for symmetric additive noise. If noise is state-dependent, sometimes jumping increases, sometimes it doesn't. In general, noise is state-dependent, so this counterexample is far from pathological: one can expect complicated behavior in most realistic situations!

2. Ball's path on landscape doesn't necessarily match transition path.

True for symmetric additive noise (in one dimension?). If noise is state-dependent, generally not true. Sometimes there can even be pretty significant deviations between the two.

This is why it is important to clearly distinguish between global landscapes (which provide global relative stability information) and local landscapes (which provide local transition path information): in general, both kinds of information are not compatible.

Just like physical motion due to a potential does not in general correspond to geodesics on that potential's surface, the shortest path between two points in our stochastic system does not in general correspond to geodesics on the landscape.

What factors determine the landscape?

Did the landscape change, or did the ball move?

Suppose that our cell is nestled comfortably in a deep valley, and that there is not enough intrinsic noise for the cell to move out of this valley in any reasonable length of time. We can perturb the cell by doing any number of things: for example, we can apply a drug, or adjust what the cell is being fed. If we manage to successfully kick the cell out of this deep valley via an external perturbation, how should we think about what we have done?

In one view, the landscape *did not change*, and the effect of our perturbation was simply to change the cell's epigenetic state (perhaps by adjusting the levels of important proteins up or down). In another view, the effect of our perturbation was to *remodel the landscape*, and the movement of our ball is just a consequence of that valley being destabilized. Which view is more correct?

Is there just one landscape, or are there many?

Some researchers have speculated that it is possible to define *the* landscape for a given biological system: a complicated surface that, if sufficiently analyzed, could in principle answer all questions regarding that system's gene expression dynamics. In particular, it should be able to answer the following questions:

1. What cell type is this iPSC most likely to eventually become? How much more likely is it to become cell type A than cell type B?

2. Which way is the ball most likely to roll through the landscape? In other words, what will its *transition path* through the landscape look like?

It turns out that, in general, it is not possible to answer both questions with the same landscape. The reason is somewhat mundane: answering questions about the probability a cell *eventually* becomes one cell type or another corresponds to asking about the steady state probability distribution, while answering questions about cell state transitions that happen in a *finite* amount of time corresponds to asking about finite time transition probability distributions. In general, these distributions can look very different.

To make this clearer, let us consider a specific example: imagine the one-dimensional system with five attractors ...[FINISH EXAMPLE LATER, OR MAYBE PUT IN A DIFFERENT SECTION]

With that, we have made the point that there are ‘global’ landscapes, which contain global relative stability information (“Which cell type is my iPSC *eventually* most likely to become?”), and ‘local’ landscapes, which contain local cell state transition information (“Which way is the cell most likely to go in the next minute or hour?”). Are these landscapes unique?

Neither global nor local landscapes are unique. Both kinds of landscapes can be freely ‘stretched’ or ‘compressed’[SEE FIG] so long as the *relative* height of each point relative to all others remains the same.

In light of the previous fact, it is probably not surprising that landscapes do not uniquely determine the underlying dynamics either—in other words, it is possible for different systems to have the same landscape. As a specific example, consider a 1D system governed by a potential V with additive noise, so that

$$\dot{x} = -V'(x) + \sigma\eta(t) .$$

The steady state distribution is easy to calculate exactly in this case (if one makes the reasonable and biological relevant assumption that P_{ss} and its derivatives vanish at infinity; see APPENDIX for more details); it is

$$P_{ss}(x) = N \exp \left(-\frac{V(x)}{\sigma^2/2} \right)$$

where N is a normalization constant. But notice that the system with dynamics governed by

$$\dot{x} = -kV'(x) + \sqrt{k}\sigma\eta(t) .$$

for some constant $k > 0$ will have the exact same distribution, since the factors of k will cancel.

Put differently, making the hills and valleys of our potential bigger or smaller will not change the landscape, so long as the noise experienced by the cell is scaled in a similar way.

Landscape properties

Generic properties of all landscapes

- Should be agnostic to the underlying kind of model (whether SDE or CME or something else). In other words, we should be able to construct a landscape whether the

- system is discrete, or continuous, or some combination.
- One minimum/valley per each stable state/attractor/cell type. This is hard to guarantee mathematically in general, but is often borne out in applications.
- Should assign each state in state space a height. In other words, the landscape is a function $f: S \rightarrow \mathbb{R}$.
- Should be bounded from below (i.e. there are no states which are impossible to escape). Along with the above point, suggests we can limit ourselves to functions $f: S \rightarrow \mathbb{R}_+$ since we can always shift the minimum height to 0.
- Should exist mathematically in fairly general circumstances (Yuan, Tang, Ao claim that P_{ss} doesn't exist in general, limiting its utility as a landscape; is this true for CME-derived systems, or only for ad-hoc systems?)
- Ideally, should involve few or none arbitrary choices; otherwise, those arbitrary choices shouldn't effect qualitative characteristics
- Need not be continuous; can be discrete (simplest example: Markov chain model)

Global vs local landscapes

It's true that, given enough time, a cell is overwhelmingly likely to go *way* over there. But what if we mainly care about where it will go in the next minute, or the next hour?

As you might imagine, where the cell is more or less likely to go strongly depends on the timescale we are interested in. For example, consider the system depicted in FIGURE: there are five attractors, each one deeper than the previous one. If a cell starts in the leftmost attractor, it might spend its first few minutes around the first two attractors. But, in the next hour, it is overwhelmingly likely to reach the third attractor. In the next 10 hours, it might be overwhelmingly likely to hit the fourth attractor. And so on.

Things a landscape can capture (but does not necessarily capture)

- Can offer estimate of probability of transition between two stable cell types (not true for local quasipotential, except if quasipotential is defined wrt one of the stable cell types in question)
- Can say something about the path of transitions between two cell types (not true for global quasipotential/not strictly true for Wang's P_{ss})
- Make the point that it might not be possible for ANY landscape to incorporate accurate transition path information and steady state relative occupancy information at the same time; the landscape you want to construct may just depend on the question you are asking.

- Two definitions given, since there are (broadly speaking) two types of landscapes: local and global
- Local landscape: provides accurate transition path information (but not relative stability information), satisfies certain other properties
- Global landscape: provides accurate relative stability information (but not transition path information), satisfies certain other properties
- Will describe below, for each landscape construction, how they satisfy our definition.

Mathematical definition

This section is intended for landscape experts comfortable with some of the mathematics. Feel free to skip this section: all of its important qualitative content will be explained in the next section through a series of examples.

Intuition for mathematical definition

Different landscapes

Simplest model: Markov chain

- Roughly agrees with the potential (possibly the potential integrated over some region of state space) for equilibrium systems. - Each stable cell type is a node. Each possible (reasonable) directed transition between stable cell types is represented by an arrow, and has an associated rate. (Some transitions may be so rare that we just suppose they are impossible and don't draw the arrow) - Generally steady state exists: landscape height is the probability of occupancy at steady state. - This kind of landscape indicates the likelihood of transitions between stable cell types, as well as relative occupancy at steady state. (for an even simpler model, just assign each stable cell type with the ss probability of occupancy; in this case, no transition information is included) - Problem: this model is particularly coarse. Reasonable to generate from empirical data, but probably hard to generate from DE/PDE models. - Most popular kind of landscape used in practice, although this discrete kind of model is rarely identified as distinct from the other types of landscapes. I think it's good to point it out explicitly.

Steady-state Fokker-Planck landscape

Wang 2008 paper (first paper discussing this?):²

Local and global quasipotential landscapes

Big 2016 review, good discussion of local/global quasipotentials:³

Vector decomposition landscapes

Sui Huang 2012 RSIF paper:⁴

Applications

Landscapes in reprogramming

Landscapes in developmental biology

Landscapes in cancer biology

Pitfalls, issues, and directions for future research

Transition path predictions

Baez stochastic mechanics⁵

Weber stochastic path integrals⁶

Wang's Waddington landscape transition path stuff⁷

Dimensional reduction

Sparseness of available data

Computational burden

Conclusion

Acknowledgements

This class file was developed by Sunrise Setting Ltd, Brixham, Devon, UK.

Website: <http://www.sunrise-setting.co.uk>

References

1. Gillespie DT. Chemical Langevin equation. *Journal of Chemical Physics* 2000; 113(1): 297–306. DOI:10.1063/1.481811.
2. Wang J, Xu L and Wang E. Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. *Proceedings of the National Academy of Sciences* 2008; 105(34): 12271–12276. DOI:10.1073/pnas.0800579105. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.0800579105>. arXiv: 1408.1149.
3. Zhou P and Li T. Construction of the landscape for multi-stable systems: Potential landscape, quasi-potential, A-type integral and beyond. *Journal of Chemical Physics* 2016; 144(9). DOI: 10.1063/1.4943096. 1511.02088.
4. Zhou JX, Aliyu DS, Aurell E et al. Quasi-potential landscape in complex multi-stable systems. *Journal of the Royal Society Interface* 2012; 9(77): 3539–3553. DOI:10.1098/rsif.2012.0434. 1206.2311.
5. Baez JC and Biamonte J. Quantum Techniques for Stochastic Mechanics. *ArXiv e-prints* 2012; 1209.3632.
6. Weber MF and Frey E. Master equations and the theory of stochastic path integrals 2016; DOI: 10.1088/1361-6633/aa5ae2. URL <http://arxiv.org/abs/1609.02849>{%}0Ahttp://dx.doi.org/10.1088/1361-6633/aa5ae2. 1609.02849.
7. Wang J, Zhang K, Xu L et al. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proceedings of the National Academy of Sciences* 2011; 108(20): 8257–8262. DOI:10.1073/pnas.1017017108. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1017017108>.