Selection Driven Scaling Regimes in Assembly Theory

Gage Siebert[1], Dániel Czégal[2], Leroy Cronin[3], and Sara Walker[1,2]

[1]School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA
[2]BEYOND Center for Fundamental Concepts in Science, Arizona State University, Tempe, AZ, USA
[3]School of Chemistry, University of Glasgow, Glasgow, UK
sara.i.walker@asu.edu

Abstract

Is there any limit to the complexity of objects that an abiotic process can construct in abundance? This question is of importance to biosignature science, but a central challenge has been finding a physically meaningful measure of complexity that can be measured in the lab. Recently, the Assembly Index defined as the smallest number of recursive joining steps to assemble an object, has been shown to be experimentally measurable for molecules. The assembly index along with copy number of objects form the foundations of observables in Assembly Theory [Sharma et al, 2023], which aims to quantify how much selection was necessary to generate a given configuration of objects. Applied to life detection, assembly theory was empirically demonstrated to distinguish chemical products derived from biological and abiotic samples [Marshall, et al. 2021]. Though the empirical results seem to place an upper bound on abiotic complexity, it is not yet possible to generalize from these measurements to environments beyond Earth without an explanatory model. Here we present an approach for calculating an object's assembly path length distribution, where an assembly path refers to a minimal sequence of assembly steps which build an object, and the expected path length is the mean of the distribution. We show, in the absence of any constraints, the expected path length scales exponentially with the assembly index. This allows us to describe the existence of two scaling regimes, one where expected path length scales exponentially with assembly index, and with sufficient constraints to lead to a linear scaling. An abrupt transition between the two regimes would be indicative of a selection-mediated phase transition.

Introduction

How can we identify life elsewhere? The past few decades have seen a focus on conceptualizing features exclusive to life - or biosignatures - especially as missions set up to look for life in the Solar System [Des Marais et al. 2002, Schwieterman et al. 2018]. However, most remote detection methods are mired in skepticism because of the difficulty of ruling out all potential abiotic explanations, and because we lack an understanding of what life could be [Vickers et al. 2023, Smith and Mathis, 2024]. This has motivated a recent shift towards complexity as an agnostic framework for life detection [Marshall et al. 2017, Johnson et al. 2018, Walker et al. 2018, Barlett et al. 2022]. Indeed, technosignatures can be understood to be the high complexity regime of biosignatures [Wright et al. 2022]. It is natural to ask, "how complex is complex enough to be a biosignature?" A boundary, if it indeed exists, should be set by the maximum complexity an abiotic process can generate [Marshall et al. 2022]. Recent progress in identifying such a boundary has been made with AT applied to chemical samples [Marshall et al. 2021]. This has been done with the complexity measure, **Assembly Index**, a, defined for an object as the fewest number of recursive, binary joining operations that can construct the object from some set of basic building blocks. When the objects of interest are molecules, the basic building blocks are generally taken to be chemical bonds, and a can be estimated empirically via nuclear magnetic resonance (NMR), tandem mass spectrometry, and infrared spectroscopy (IR) [Jirasek et al. 2023]. Empirical tests of samples taken from diverse environments, including blinded samples, have shown strong evidence for a limit to the a of abiotic molecules, suggesting that the boundary for biosignature complexity of terrestrial samples is no lower than $a = 15$ [Marshall et al. 2021]. To confidently extend these results beyond our biosphere, we require an explanation for this observation that generalizes across diverse planetary environments. Given that a meaningful biosignature captures the necessary features of an observation of life, confidently claiming what constitutes a biosignature also necessarily implies a profound statement about what life is [Tirard et al. 2010]. The theory we develop and report herein identifies the physical phenomena of life with open ended exploration of possible assembled objects, which could be of use for identifying the origin of life in the lab [Asche et al. 2021], new forms of artificial life [Corominas-Mutra et al. 2018], and in life detection beyond Earth [Walker et al. 2018].

It has remained difficult to formalize the seemingly straightforward observation that abiotic and biotic systems display qualitatively different complexification behavior and tend to remain abiotic or biotic, respectively. Complex abiotic systems, in accordance with what one would expect from the second law of thermodynamics, have a tendency toward decay, simplification and forming tars [Benner et al. 2010]. Biotic systems not only maintain their configuration stably over time, but can and often do, explore new complex configurations over time [Darwin, 1859] in an open-ended way [Soros and Stanley 2014, Corominas-Mutra et al. 2018]. Not only has it proven difficult to engineer an origin of life event [Preiner et al. 2020], we only know of one such event in the observable universe, confounding our ability to generalize statistics of its likelihood

[Carter 1983]. Many lines of evidence are consistent with the boundary between abiotic and biotic systems being an unstable equilibrium in terms of complexification dynamics, suggestive of two dynamical phases of matter: abiotic and biotic.

A priori, AT puts all joining operations on the same footing to study the combinatorial nature of complex object construction. Selection is then most simply described as a divergence from assigning uniform probabilities to all joining operations as introduced by Sharma et al. [Sharma et al, 2023]. Consider the dynamics of a finite system which applies joining operations with no selection. Since the number of unique objects constructable after k operations is at least exponential in k, the expected copy number as a function of k will quickly approach 1. Sharma et al., referred to such systems as undirected, and their behavior is typical of complex abiotic systems (ex: large polymer construction in abiotic tar). This already suggests a mechanistic explanation of the observed maximum a of abundant abiotic molecules. In any undirected and therefore abiotic system, either complex molecules are not forming, or material diffuses so widely over the space of possible molecules that the odds of the system constructing two identical molecules with large a is vanishingly small. Biotic chemical systems are not well described by undirected assembly; they form many highly complex (large a) molecules in large copy numbers, which requires selection. This distinction between directed and undirected assembly dynamics suggests a phase transition.

Herein, we take steps towards understanding whether a phase transition is indeed the right framing to distinguish systems that cannot complexify and those that do in an open-ended way, using the formalism of AT. The sharpness of the empirical $a \approx 15$ boundary observed by Marshall et al. suggests an abrupt phase transition in assembly dynamics [Marshall et al. 2021]. A common indication of a phase transition is the presence of regions of phase space with qualitatively distinct scaling behaviors [Goldenfeld 1992]. To explore this further, we develop an algorithm for calculating the distribution of assembly path lengths to construct a given object. We then introduce constraints on this distribution, parameterizing the probability that a joining operation can be added to an assembly path. Expected assembly path length then becomes a function of these constraints. We find two scaling regimes: when the applied constraints are low, expected assembly path length is exponential in object a, and when they are high, expected assembly path length is linear in object a. This result suggests that selection is required to construct an object in a-linear steps, and in the absence of selection there is a rapid divergence in the number of steps required to construct complex objects.

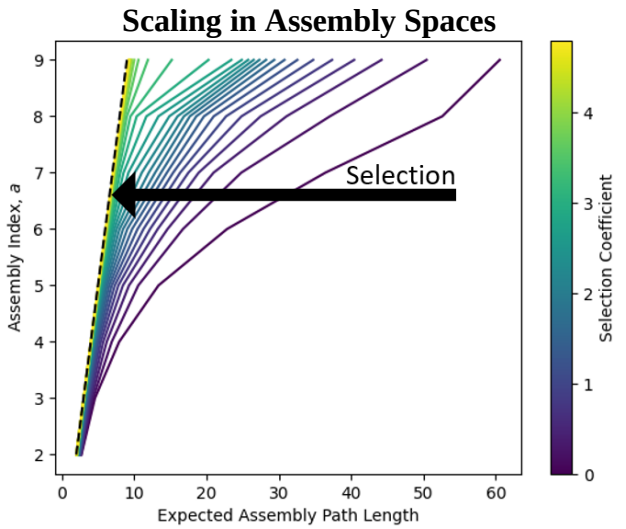

Figure 1: Shown are the expected path length of integers, where assembly paths are addition chains, as a function of a parameterized selection coefficient. Here we average the results of many objects of a given a together. In the low selection limit, the length of a typical assembly path is exponential in a, while in the high selection limit it grows linearly with a.

To make statements about the expected path length, we calculated the normalized distribution of all possibly assembly paths to construct an object. The number of paths for even relatively low a objects is astronomically large, so our algorithm is designed to effectively sample from the space of all paths to generate estimates of assembly path lengths. Though we are primarily interested in molecules, this algorithm generalizes to other object classes like addition chains and strings, which are important for artificial life applications where the substrates of life could be vastly different [Lenski et al. 2003, Kempes and Krakauer 2020], Figure 1. For complex molecules, we find the expected path length scales linearly with the size of the molecule (number of bonds), but exponentially with a. In each assembly space we have analyzed, the expected path length scales exponentially in a in the absence of constraints, however with sufficient constraint, it scales linearly in a. The different scaling regimes in the low and high constraints regions of parameter space are consistent with the presence of an abiotic-to-biotic phase transition driven by selection, in agreement with previous considerations [Sharma et al. 2023] and observations [Marshall et al. 2021] and should inform future models aiming to characterize a phase transition from abiotic to biotic across a variety of substrates.

Acknowledgements

Authors acknowledge support from John Templeton Foundation award number G09914-300 and from NSF award number 2233001.

References

Asche S., Cooper G. J. T., Keenan G., Mathis C., Cronin L. (2021). A robotic prebiotic chemist probes long term reactions of complexifying mixtures. *Nat. Commun*., 12(3547).

Barlett S., Li J., Gu L., Sinapayen L., Fan S., Natraj V., Jiang J. H., Crisp D., Yung Y. L. (2022). Assessing planetary complexity and potential agnostic biosignatures using epsilon machines. *Nature Astronomy*, 6:387-392.

Benner S. A., Kim H. J., Kim M. J., Ricardo A. (2010). Planetary organic chemistry and the origins of biomolecules. *Cold Spring Harb Perspect Biol*. 2(7):a003467

Carter B. (1983). The anthropic principle and its implications for biological evolution. *Phil. Trans. R. Soc. A.*, 310:347-363.

Corominas-Mutra B., Seoane L. F., Solé R. (2018). Zipf's Law, unbounded complexity and open-ended evolution. *J. R. Soc. Interface*, 15:20180395.

Darwin, C. (1859). *The Origin of Species*. John Murray, London, UK.

Des Marais, D. J., Harwit, M. O., Jucks, K. W., Kasting, J. F., Lin, D. N. C., Lunine, J. I., Schneider, J., Seager, S., Traub W. A., Woolf, N. J. (2002) Remote Sensing of Planetary Properties and Biosignatures on Extrasolar Terrestrial Planets. *Astrobiology*, 2(2):153-181.

Goldenfeld N. (1992). Lectures On Phase Transitions And The Renormalization Group. Westview Press, Boulder, CO.

Jirasek M., Sharma A., Bame J. R., Mehr S. H. M., Bell N., Marshall S. M., Mathis C., Macleod A., Cooper G. J. T., Swart M., Mollfulleda R., Cronin L. (2023). Determining Molecular Complexity using Assembly Theory and Spectroscopy. *Arxiv*, preprint.

Johnson, S. S., Anslyn E. V., Graham H. V., Mahaffy P. R., Ellington A. D. (2018). Fingerprinting Non-Terran Biosignatures. *Astrobiology*, 18(7): 915-922.

Kempes C. P., Krakauer D. C. (2020). The Multiple Paths to Multiple Life. *J. Mol. Evol.*, 89:415-426.

Lenski R. E., Ofria C., Pennock R. T., Adami C. (2003). The evolutionary origin of complex features. *Nature*, 423:139-144.

Liu Y., Mathis C., Bajczyk M. D., Marshall S. M., Wilbraham L., Cronin L. (2021). Exploring and mapping chemical space with molecular assembly trees. *Sci. Adv.*, 7(39):eabj2465.

Marshall S. M., Mathis C., Carrick E., Keenan G., Cooper G. J. T., Graham H., Craven M., Gromski P. S., Moore D. G., Walker S. I., Cronin L. (2021). Identifying molecules as biosignatures with assembly theory and mass spectrometry. *Nat. Commun.*, 12(3033).

Marshall S. M., Moore D. G., Murray A. R. G., Walker S. I., Cronin L. (2022). Formalizing the Pathways to Life Using Assembly Spaces. *Entropy*, 24(7):884.

Marshall S. M., Murray A. R. G., Cronin L. (2017). A probabilistic framework for identifying biosignatures using Pathway Complexity. *Phil. Trans. R. Soc. A.*, 375(2109).

Preiner M., Asche S., Becker S., Betts H. C., Boniface A., Camprubi E., Chandru K., Erastova V., Garg S. G., Khawaja N., Kostyrka G., Machné R., Moggioli G., Muchowska K. B., Neukirchen S., Peter B., Pichlhöfer E., Radványi Á., Rossetto D., Salditt A., Schmelling N. M., Sousa F. L., Tria F. D. K., Vörös D., Xavier J. C. (2020). The Future of Origin of Life Research: Bridging Decades-Old Divisions. *Life*, 10(3):20.

Schwieterman E. W., Kiang N. Y., Parenteau M. N., Harman C. E., DasSarma S., Fisher T. M., Arney G. N., Hartnett H. E., Reinhard C.T., Olson S. L., Meadows V. S., Cockell C. S., Walker S. I., Grenfell J. G., Hedge S., Rugheimer S., Hu R., Lyons T. W. (2018). Exoplanet Biosignatures: A Review of Remotely Detectable Signs of Life. *Astrobiology*, 18(6):663-708.

Sharma A., Czégel, D., Lachmann M., Kempes C. P., Walker S. I., Cronin L. (2023). Assembly theory explains and quantifies selection and evolution. *Nature*, 622:321-328.

Smith H. B., Mathis C. (2023). Life detection in a universe of false positives. *BioEssays*, 45(2):2300050.

Soros L. B., Stanley K. O. (2014). Identifying Necessary Conditions for Open-Ended Evolution through the Artificial Life World of Chromaria. *Artificial Life XIV*. MIT Press, Cambridge, MA.

Tirard S., Morange M., Lazcano A. (2010). The Definition of Life: A Brief History of an Elusive Scientific Endeavor. *Astrobiology*, 10(10):1003-1009.

Vickers P., Cowie C., Dick S. J., Gillen C., Jeancolas C., Rothschild L. J., McMahon S. (2023). Confidence of Life Detection: The Problem of Unconceived Alternatives. *Astrobiology*, 23(11):1202-1212.

Walker S. I., Bains W., Cronin L., DasSarma S., Danielacho S., Domagal-Goldman S., Kacar B., Kiang N. Y., Lenardic A., Reinhard C. T., Moore W., Schwieterman E. W., Shkolnik E. L., Smith H. B. (2018). Exoplanet Biosignatures: Future Directions. *Astrobiology*, 18(6):779-824.

Wright J. T., Haqq-Misra J., Frank A., Kopparapu R., Lingam M., Sheikh S. Z. (2022). The Case for Technosignatures: Why They May Be Abundant, Long-lived, Highly Detectable, and Unambiguous. *ApJL*, 927(2):L30.