

# Quantifying the origins of life on a planetary scale

Caleb Scharf<sup>a,1</sup> and Leroy Cronin<sup>b,1</sup>

<sup>a</sup>Columbia Astrobiology Center, Columbia Astrophysics Laboratory, New York, NY 10027; and <sup>b</sup>School of Chemistry, University of Glasgow, Glasgow, G12 8QQ, United Kingdom

Edited by Neta A. Bahcall, Princeton University, Princeton, NJ, and approved May 17, 2016 (received for review November 23, 2015)

**A simple, heuristic formula with parallels to the Drake Equation is introduced to help focus discussion on open questions for the origins of life in a planetary context. This approach indicates a number of areas where quantitative progress can be made on parameter estimation for determining origins of life probabilities, based on constraints from Bayesian approaches. We discuss a variety of “microscale” factors and their role in determining “macroscale” abiogenesis probabilities on suitable planets. We also propose that impact ejecta exchange between planets with parallel chemistries and chemical evolution could in principle amplify the development of molecular complexity and abiogenesis probabilities. This amplification could be very significant, and both bias our conclusions about abiogenesis probabilities based on the Earth and provide a major source of variance in the probability of life arising in planetary systems. We use our heuristic formula to suggest a number of observational routes for improving constraints on origins of life probabilities.**

origin of life | planetary scale | chemical search space | exoplanetary science

The question of whether or not life exists beyond the confines of the Earth environment is intimately related to the question of life's origins on Earth or anywhere else. For example, if just one instance of life with an independent origin is detected—whether in the solar system or on a suitable exoplanet—estimates of the Bayesian probability for life across the universe will be significantly improved. Specifically, the rate of abiogenesis per young planet could, in principle, be constrained to be at least one event per Gyr (1). Equally, if a detailed mechanism for the terrestrial origins of life (OoL) were identified and tested, a complementary estimate of the rate of planetary abiogenesis should be possible. Furthermore, the architecture of a given solar system might also lead to widely different probabilities (e.g., the proximity of Mars and Earth and their potential exchange of material might result in a fundamentally larger search space in which OoL events could occur). Discovering complex organic molecules on Mars could help elucidate this possibility (2).

At present, OoL science involves a wide range of approaches, theories, and opinions. A considerable body of work now exists on hypothetical early terrestrial life and its precursor chemistry including the RNA-world hypothesis, prior polymerization chemistry, autocatalytic processes, and metabolic origins, as well as the presumed transition to fully cellular organisms. Significant work has also been undertaken on inorganic “template” chemistry, the planetary evolution of suitable material precursors, and important substrates for the emergence of a biology—including the study of electrochemical gradients in hydrothermal vents systems and processes such as serpentinization as energy sources. In addition, the genomic exploration of the modern Earth's microbial life is providing ever-deeper insight into the nature and evolution of the biosphere's nested metabolic processes (3), which may provide clues to the original energy landscape of prebiotic chemical cycles. Theoretical and computational work on emergent and dynamical systems, complexity, and “dry” (not wet laboratory-based) artificial life (A-Life) are also contributing to our picture of the fundamental principles of biological system operations.

However, there is no strong consensus that supports any single OoL hypothesis or timeline—whether for the specific OoL events on the Earth or for any plausible alternatives. It is therefore

currently impossible to construct a first principles quantitative estimate of an abiogenesis probability for the young Earth. It is also the case that we typically conflate the idea of “microabiogenesis”—namely, a highly localized assembly of molecular structures that meets the minimum criterion for a living system—with “macro-” or planet-wide abiogenesis. In other words the idea of life arising on a planet is treated as a singular event rather than a possible accumulation of many microscale events, each of which might be considered an origin event in its own right, and are arguably more accessible through laboratory or modeling investigations. Nonetheless, abiogenesis did clearly occur at least once on Earth (4), and we do possess general information about Earth's composition and a number of notional histories for the surface and near-surface environment during the planet's first few hundred million years following the solar system's initial collapse and condensation from the presolar nebula.

## Results and Discussion

In this short paper, a simple formula is proposed to specifically encourage further consideration of the issues at play in linking OoL (or abiogenesis) to planetary environments. This approach parallels similar descriptive and pedagogical methods applied in assessing the likelihood of extraterrestrial communications in the Search for Extraterrestrial Intelligence (SETI). A simple application of this formula is made, and the results are discussed below. Furthermore, a number of modifications and extensions are considered that should help bring focus to the regions of parameter space where tractable progress can be made on building a genuinely quantitative evaluation of OoL likelihoods (5).

In 1961, Frank Drake introduced an equation to illustrate the factors involved in estimating the potential number of communicative civilizations in the Milky Way galaxy. This formula, now generally known as the Drake Equation, has served as a useful tool for focusing discussion on the extent to which we do, or do not, have constraints on its various factors and for stimulating ideas

## Significance

**In this paper, we describe an equation to estimate the frequency of planetary “origin of life”-type events that is similar in intent to the Drake Equation but with some key advantages—specifically, our formulation makes an explicit connection between “global” rates for life arising and granular information about a planet. Our approach indicates scenarios where a shared chemical search space with more complex building blocks could be the critical difference between cosmic environments where life is potentially more or less abundant but, more importantly, points to constraints on the search. The possibility of chemical search-space amplification could be a major variance factor in planetary abiogenesis probabilities.**

Author contributions: C.S. and L.C. performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

<sup>1</sup>To whom correspondence may be addressed. Email: caleb@astro.columbia.edu or lee.cronin@glasgow.ac.uk.

**Table 1. Plausible values for the factors in the OoL frequency equation for the Earth**

Factor	Value	Basis for value
$N_b$	$\sim 10^{49}$	No. of atoms in Earth's modern crust + ocean + atmosphere
$\langle n_o \rangle$	$\sim 10^{11}$	No. of atoms in bacterium (dry weight)
$f_c$	$\sim 10^{-14}$	Fraction based on the ratio of modern Earth's total estimated biomass to the total mass of crust + ocean + atmosphere
$P_a$	$\sim 10^{-36}$	Assuming $\lambda = 10^{-3}$ per Gyr
	$\sim 10^{-30}$	Assuming $\lambda = 10^3$ per Gyr
	$\sim 10^{-33}$	Assuming $\lambda < 1$ per Gyr (12% chance)

about how to make scientific progress on the problem of finding life in the universe over a given period (6). There are however serious limitations to actually using the Drake formula to produce meaningful estimates of life's frequency in the universe. A particular issue is that although the formula seeks to describe the entire galaxy's living population, it does not explicitly allow for the possibility of life spreading and expanding between planets. Thus, the factors in the equation may *not* be independent at all.

Here, we propose that a focus on the detailed parameters involved in planetary OoL offers a better-constrained entry point to this type of estimation and could produce practical results, especially when combined with the search for potentially "habitable" exoplanetary systems. Assuming that life on Earth did originate in situ—in other words, that the transition, or the key transitions, between a nonliving state and a living state occurred within the planetary environment and not from far off-world (e.g., panspermia beyond the planetary system)—it should be possible to construct a high-level quantitative description of the basic features of an OoL, irrespective of the details. Such a description necessarily makes a quantitative connection to planetary properties, such as mass, element abundance, and energy resources.

The following OoL frequency equation is therefore proposed for  $\langle N_{\text{abiogenesis}}(t) \rangle$ , the *mean* expected number of a planet's abiogenesis (or origin) events, in time  $t$  (7) (i.e., the mean over a number of equivalent instances), where the definition of such events and the associated factors are discussed in the subsections below:

$$\langle N_{\text{abiogenesis}}(t) \rangle = N_b \cdot \frac{1}{\langle n_o \rangle} \cdot f_c \cdot P_a \cdot t. \quad [1]$$

As discussed in more detail below, the definition of abiogenesis probability ( $P_a$ ) is different from that often quoted—it is not a global (planet) rate per se but a probability per unit time *per set of chemical building blocks*. This treatment also assumes for simplicity that abiogenesis behaves as a Poisson process, thus the left-hand side is an estimated average quantity. This approach allows us to begin to unpack the specific planetary properties that may influence net abiogenesis rates.

**$N_b$ : Number of Potential Building Blocks.** Modern terrestrial life can be deconstructed into a large, but finite, set of functional chemical components. For example, the major building blocks are often cited as consisting of families of carbohydrates, lipids, proteins, and nucleic acids. Additional blocks might include metal ions as part of the metallome (e.g., coordinated zinc ions stabilizing protein fold structures, iron in hemoglobin, copper in hemocyanin, and other metalloproteins), silicon (e.g., in plants), and many trace elements.

However, life today does not necessarily reflect all of the details of early biochemistry, or earlier complex chemistry that was able to self-sustain routes to competing biologies that ultimately led to biology as we know it. Life today is more likely to reflect the consequence of billions of years of evolution of a series of earlier chemical toolkits, all with decreasing complexity (at earlier stages) and, therefore, increasing probability of emergence over a short period. In that context, it might be advisable to break these components down further and consider the raw elemental (atomic) abundances as the first factor in the OoL equation. An additional advantage in doing so is that element abundance simplifies the connection of  $N_b$  to bulk planetary properties, thereby linking this expression back to the scope of the Drake Equation and our direct physical knowledge of the Earth and other planets.

In this case,  $N_b$  represents a maximal set of building blocks for life that can be estimated from the total mass and composition of the outer planetary layers (see below). In the Earth's case, it is self-evident that not all of the atoms of Earth's outer layers are incorporated into living states, nor are all available for life at any given time. To account for element availability, the availability factor,  $f_c$ , is described further below. Also, any collection of building blocks for life must be able to interact with each other and provide routes by which a subset of building blocks can drive the assembly and complexification of a subset of new building blocks—that have improved function, structure, and robustness. Chemical cycles, operating on these building blocks will provide raw materials as well as facilitate the assembly of more sophisticated building blocks from the minimal set, but these cycles are fundamentally dependent upon the input building block set.

Chemical cycles could also allow the same building blocks, in different sequences, to make polymers and other infrastructure that can allow both control of assembly and maintenance of sequence and expression of function. Also, the kinetic connections between the different cycles and sequences constitutes a minimal kinetic-genetic mechanism by which a more robust and ultimately evolvable genetic machinery may develop, and such machinery might even use an network of so-called "inorganic" reactions to start developing function (8).

**$\langle n_o \rangle$ : Mean Number of Building Blocks per "Organism," or Biochemically Significant System.** The definition of "organism" in this context is broad and is related to a quantitative definition of life—the precise details of which remain an open question at this time. For the OoL, it could be taken to mean a minimal lifeform, one that is capable of homeostasis, reproduction, and open-ended evolution. However, we might determine that such a lifeform exists either as an encapsulated, cellular system or as a distributed (nonlocal) but interdependent and autocatalytic chemical system. Similarly, a biochemically significant system could be construed to be a system or environment that is a direct and necessary precursor to a minimal lifeform, with an extremely high probability of the transition to that lifeform. Properly evaluating  $\langle n_o \rangle$  may therefore require the development of a coherent and quantitative definition of the "aliveness" of a collection of building blocks. In addition, the gradual development of entities that are more complex, but not quite "alive," represents a complication in that  $\langle n_o \rangle$  might not be a simple threshold.

**$f_c$ : Fractional Availability of Building Blocks During Time  $t$ .** Within any given timespan,  $t$ , it can be assumed that only a certain fraction of the total number of potential building blocks in a planetary environment will actually be available for life. Availability in this context may be defined through a number of factors: "free" (unbound to other molecular or atomic species), "mobile" (capable of physical transport, not restricted except in terms of necessary localization), or "energetic" (energetically favored chemical bonding and incorporation into a system). These factors can be strongly dependent on planetary details, such as environmental

temperature. The use of this factor (i.e., not incorporating it directly into  $N_b$ ) allows the weighting given to intrinsic planetary properties such as crust/mantle ratios or global elemental abundances to be separated and adjusted independently. As we discuss below,  $f_c$  can be significantly expanded to include explicit treatments of these details of planetary habitability and the interplay between living systems and environment (e.g., the sequestration of building blocks by prior life).

**$P_a$ : Probability of Assembly per Unit Time.** This factor refers to the probability of an abiogenesis “event” per unit time, per set of suitable building blocks, through the assembly of those building blocks. The necessary definition of an event is of course very poorly understood. Abiogenesis could be highly spatially localized, within a free-floating vesicular structure for example, or in an inorganic substrate’s intimate topography. Alternatively, abiogenesis could refer to a nonlocalized, sequential series of assembly steps or processes that lead directly and irreversibly to an evolvable system. We purposefully treat  $P_a$  as a catchall that circumvents the need to go into the (unknown) mechanical details.

This definition of  $P_a$  is distinct from that often considered. Ref. 1 defines a parameter  $\lambda$  as a probability of life arising per unit time (as a Poisson process). However, this refers to a net global rate (rate per planet), whereas we propose that the root probability of interest is that which refers to the probability per unit time per suitable set of building blocks—therefore introducing factors directly related to the planetary properties themselves. Thus, the relationship between the global rate and  $P_a$  is:

$$\lambda = N_b \cdot \frac{1}{\langle n_o \rangle} \cdot f_c \cdot P_a. \quad [2]$$

Critically,  $P_a$  could, in principle, be determined “in vitro (or conceivably “in silico”), whereas  $\lambda$  is unlikely to be determined ab initio but, instead, may be constrained from astrobiological investigations of biosignatures elsewhere in the solar system or in exoplanetary systems. For example, if experiments investigating potential prebiotic chemical networks succeed in enabling the emergence of just one functioning, evolvable chemical system, empirical constraints could be made on both  $P_a$  and  $\langle n_o \rangle$  as well

as identifying the required building blocks (as elements or simple molecules) that constitute  $N_b$ .

**Evaluating  $N_{abiogenesis}$  and  $P_a$ .** The motivation for choosing the  $N_{abiogenesis}$  term on the left-hand side of the OoL frequency equation is largely one of emphasis. However, our definition is subtly different from others. In particular, it is critical that it be understood in the context of work that seeks to build a Bayesian framework for assessing the probability of life in a global fashion (1). Following the methodology of ref. 1, if (see below)  $\lambda$  is the probability per unit time of life developing (i.e., the mean number of times life develops per unit time) and abiogenesis is treated as a Poisson process (implicitly assuming that separate abiogenesis events or assembly events do not influence each other), then the probability of a given  $N_{abiogenesis}$  is

$$P[N_{abiogenesis}, \lambda] = e^{-\lambda(t-t_{min})} \frac{\{\lambda(t-t_{min})\}^{N_{abiogenesis}}}{n!}, \quad [3]$$

where, again,  $\lambda$  is the probability of  $n$  abiogenesis events on a planet per unit time and  $\lambda = N_b \cdot \frac{1}{\langle n_o \rangle} \cdot f_c \cdot P_a$ . The quantity  $t_{min}$  is a time following the formation of a planet (Earth) during which the physical conditions are definitely incompatible with life (e.g., during the period where the surface temperature was too high for any stable organic molecules to exist). Following the results of ref. 1,  $t_{min}$  is plausibly set at 0.5 Gyr for an initial terrestrial formation time 4.5 Gya. Thus, for a Poisson process,  $\langle N_{abiogenesis}(t) \rangle$  is simply the product of rate and time.

As described above, we propose the OoL frequency expression (Eq. 1) to help facilitate and focus discussion on OoL science in the same way that the Drake Equation is applied to the quest for life elsewhere in the universe. Eq. 1 is not meant to be genuinely predictive, nor complete in the form given here. Despite all of these shortcomings, it is nonetheless instructive to evaluate a range of plausible values for the factors.

As discussed in detail in ref. 1 and by earlier works (4, 7), it is possible to deduce an extremely wide range of global abiogenesis probability rates ( $\lambda$ ) (on the basis of what we currently know about life’s existence on Earth and hypotheses based on observations of the surrounding cosmos. Nonetheless, a carefully constructed Bayesian analysis (1) suggests that a range of  $10^{-3} \leq \lambda \leq 10^3 \text{Gyr}^{-1}$

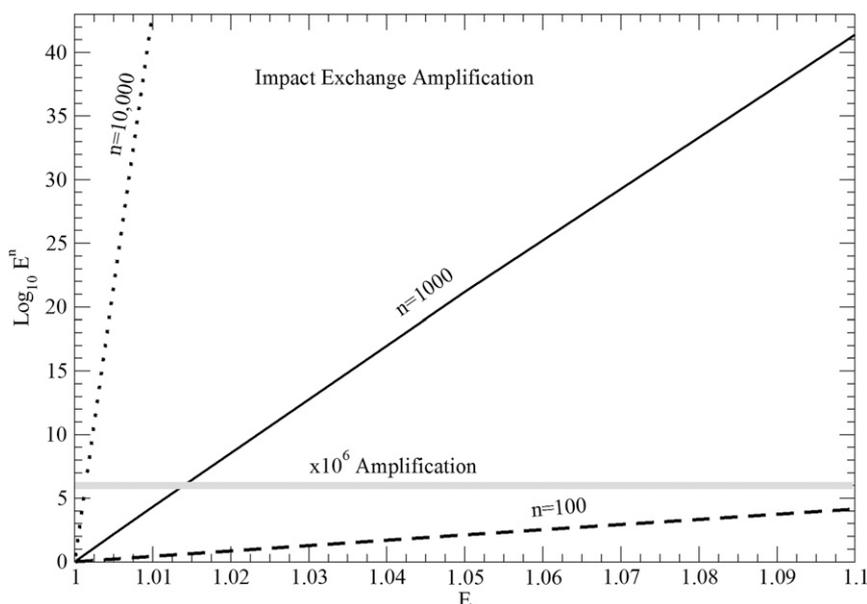


Fig. 1. Illustration of the dependence of probability amplifications ( $E^n$ ) that result from exchange histories with  $n = 100, 1,000, 10,000$ .

is allowed, with a 12% chance that  $\lambda < 1 \text{ Gyr}^{-1}$ , assuming an uninformative logarithmic prior for  $\lambda$ .

In Table 1, a set of crude but plausible values are listed for the factors of Eq. 1. The basis is a simplistic overview of the Earth. Potential building blocks are taken to be the elemental (atomic) constituents of the upper planetary environment (crustal mass, oceans, and atmosphere). Using the entire upper planetary environment is very likely a gross overestimate, but even a few orders of magnitude variation will not alter the general conclusions. The number of building blocks required for abiogenesis (see above) is derived from an approximate mean value for the number of atoms in a bacterium. The fractional availability of building blocks is derived from estimates of the modern ratio of Earth's total biomass to the mass of the upper planetary environment.

Although the chosen values are spectacularly approximate, the linear nature of the equation ensures that variations in any factors by even a few orders of magnitude do little to alter the basic outcome. In the scenarios considered, the abiogenesis probability per unit time per set of building blocks is found to range from  $P_a \sim 10^{-36}$  to  $P_a \sim 10^{-30}$ , with a value of  $P_a < 10^{-33}$  being broadly compatible with the specific case of the Earth. Thus, in this case, Eq. 1 indicates that a very small assembly probability per unit time could in principle be compensated for by the large scale of a planetary “search engine” —readily producing at least one abiogenesis event within a geologically appropriate timespan.

Obviously, our equation is heuristic in nature, and we have set the problem up so that probabilities like these are not too surprising. However, this result is independent of fine details. It does not, for example, hinge on any analysis of the combinations and permutations of atoms required to form a functional biochemical system or the odds of assembly computed on that basis. All of that complexity is folded implicitly into  $P_a$ . Furthermore, critically, Eq. 1 now enables us to examine the potential influence of planetary physical characteristics on abiogenesis likelihoods. As long as we remain cognizant of the fact that (i) we do not understand the OoL on Earth, (ii) we do not know whether the OoL on Earth is representative in any way of how the OoL could, hypothetically, occur elsewhere.

In the above, the original estimation of  $\lambda$  is agnostic to whether there is a sole abiogenesis event or more than one (i.e., the constraints assume the probability of life arising  $n \geq 1$  times in time  $t$ ). That estimation therefore does not preclude cases where  $P_a$  is actually much larger, and many independent abiogenesis events occur during a given time that are indistinguishable in our current model for the history of life on Earth (although they could, leading to the inaccurate conclusion that  $N_{abiogenesis} = 1$ ). Similarly, there could be many independent abiogenesis events that undergo rapid selection competition, or simple chemical failure, thereby winnowing the population to a single progenitor model for all of life today.

Other games can be played. For example, one could demand that  $P_a = 1$  in a 100 Myr timespan (i.e.,  $P_a = 10^{-8}$  in a year) and that in all planetary instances, there will only ever be a single abiogenesis outcome during that time,  $\langle N_{abiogenesis} \rangle = 1$ . Retaining the Table 1 values for  $N_b$  and  $f_c$ , a literal interpretation would be that  $\langle n_o \rangle \sim 10^{35}$ . In other words, an inevitable net abiogenesis in 100 Myr (via a global process of interaction and chemical selection over geological timescales) might involve a total number of potential building blocks that amounts to  $\sim 10^{24}$  times the blocks required for a single, final, viable organism. Again, we emphasize that in this initial simple, heuristic, form the equation is not meant to be rigorous. It can however drive discussion of the macro- and microscale issues for the OoL and form the basis for a more complete quantitative measure of abiogenesis rates, as described below.

**Unpacking Planetary Properties.** Eq. 1 assumes that all factors are independent and that they contribute linearly to  $\langle N_{abiogenesis} \rangle$ .

The independence of factors, of course, may not be true. For example, as building blocks (atoms or simple molecules) are incorporated into prebiotic molecular structures, the building blocks' availability will change. If polymerization is a key part of abiogenesis, this abundance dependency could be important as an asymptotic property. Similarly,  $P_a$  is unlikely to be a single, time-independent variable. Rather,  $P_a$  will have a complex dependency on many factors, from the scale of the planetary environment (and therefore  $N_b$ ) to the relative abundance of building blocks (i.e., a block dependent  $f_c$ ) and the chemical pathways involved in assembly.

The factor  $f_c$  in particular can be readily expanded out. For example:

$$f_c = f_p \cdot f_a \cdot f_e \cdot (1 - f_i), \quad [4]$$

where  $f_p$  is the fraction of the total planetary environment (containing  $N_b$ ) within the “habitable zone” (e.g., using the common astronomical definition of temperature and vapor pressure such that open water remains liquid),  $f_a$  is the fraction of building blocks in solution or part of a bioaccessible substrate,  $f_e$  is the fraction of building blocks with access to the necessary energy (chemical, thermal, photonic) to drive assembly to a biologically meaningful assembly, and  $f_i$  is the fraction of building blocks that may already be incorporated into abiogenesis related assemblies (i.e., not available at time  $t$  for incorporation in any new abiogenesis events).

In this instance,  $f_p$  is an explicit function of the planetary climate state, which is in turn determined by properties such as the planetary spin-orbit state, the parent star, and the planetary atmospheric composition.  $f_a$  is more complex but could be estimated from initial planetary veneer compositions. The factor  $f_e$  might also be estimated from models of initial planetary properties. In the simplest case, for  $t = 0$ , the factor  $f_i$  is assumed to be 0. Allowing  $f_i > 0$  introduces explicit nonlinearity in the equation because we might naively expect  $f_i = (N_i \cdot \langle n_o \rangle) / N_b$ , where  $N_i$  is the number of existing “organisms” at time  $t$  that are sequestering building blocks away from abiogenesis events (in other words we assume that the building blocks do not actually contribute positively to further abiogenesis). Thus, the full expression for abiogenesis events now becomes

$$\langle N_{abiogenesis}(t) \rangle = N_b \cdot \frac{1}{\langle n_o \rangle} \cdot f_p \cdot f_a \cdot f_e \cdot \left( 1 - \frac{N_i \langle n_o \rangle}{N_b} \right) \cdot P_a \cdot t. \quad [5]$$

As Eq. 5 is written, there is no explicit incorporation of the concept of dilution. A planet may represent a large set of  $N_b$  and high  $f_c$  as defined, but the building blocks could still be highly diluted, for example by large water bodies. Indeed, the dilution problem of an ocean-covered Hadean/Archean Earth is well appreciated (where monomer concentration in open oceans appears too low for polymerizations to occur). In this case, because  $N_b$  presumably scales with planetary size, any potential dilution factor may also scale with  $N_b$ . Thus, there must be a more complex behavior of  $f_c$ , where the “availability” of building blocks should include a statistical factor to account for localized concentration or spatial clustering (e.g., hydrothermal vent “oasis” environments, tidal zones).

The final  $N_{abiogenesis}$  after time  $t$  could also depend on a survival or failure rate. Similar to our above evaluation of Eq. 1,  $P_a$  could be larger but balanced by a probability of extinction attributable to molecular fragility, system fragility, competition, or just a steady rate of random failure.

**Planetary Exchange and Chemical Search Space.** It is possible that material exchange between terrestrial-type planetary surfaces could serve to greatly amplify the chemical “search space” within a planetary system. If we assume, for example, that the Hadean

Earth and Noachian Mars provided chemical incubators that were distinct from other solar system locales (e.g., primordial cometary or asteroidal environments) and well suited to the emergence of increasing prebiotic chemical complexity, then exchange between these bodies could have accelerated the exploration of chemical novelty.

There is good evidence that the architecture and dynamical evolution of the solar system resulted in early periods of significant impact events on both Earth and Mars and subsequent ejection and exchange of lithospheric material (9). Although the exchange of viable organisms between these planets is still very much an open question (10), the exchange of molecular species may be much more likely. Furthermore, the energy scale of impact events required to launch material onto interplanetary trajectories need not be so destructive as to “reset” indigenous prebiotic molecular complexity and populations on a planetary scale.

In this scenario,  $N_{\text{abiogenesis}}$  becomes coupled across planets in a potentially complicated, but important way. The variation in  $N_b$  and  $f_c$  attributable to impact ejecta exchange may be very small; however, changes in  $P_a$  as defined may be significant. As described above, we see  $P_a$  as the probability of an “event”—either a specific chemical assembly or a nonlocalized sequence or chemical system that leads irreversibly to an evolvable system. Thus, in the same spirit as Eq. 1, we suggest that the effect of interplanetary exchange could be most simply modeled by assuming that all receipts of material (e.g., Earth receiving Mars material) result in an increase in  $P_a$ . In other words, it is assumed that “new” molecular species are brought to the planet and that these species positively influence the pathway to abiogenesis.

For example, this positive influence could behave as a simple geometric series:

$$P'_a = P_a \cdot E^n, \quad [6]$$

where  $E$  is the factor by which  $P_a$  is increased by a single exchange,  $n$  is the total number of “receipt” events attributable to impact ejecta (e.g., the number of impact ejecta creating collisions that occur on Mars), and  $P'_a$  is the revised assembly event probability.

The factor  $E$  is of course unknown and likely varies with each exchange. However, in an extreme scenario, each planetary incubator generates an entirely unique set of molecular species between impact-exchange events. Very crudely speaking, in this case, it would be easy for  $E \sim 2$  and, as a trivial example, even with modest exchange of  $n \sim 100$  then  $P'_a \sim 10^{30} P_a$ . However, even if  $E$  is far smaller, the amplification attributable to exchange can be large. Such amplification raises an interesting number of issues. First, natural variation in abiotic material exchange between different planetary systems could introduce significant spread in the net probability of life occurring on a planet per unit time ( $\lambda$ ), thereby biasing inferences about cosmic probabilities (1). Second, if significant exchange amplification occurred in the solar system, the amplification might obscure the baseline value of  $P_a$  for either Earth or Mars.

The value of  $n$  can in principle be constrained in the solar system by data on impact histories and our models of solar system dynamical evolution. For the Earth, evidence suggests that the Late Heavy Bombardment (thought to occur between 3.8 and 4.1 Gya) may have involved more than 20,000 Earth impacts capable of launching ejecta to interplanetary space (11). Mars appears likely to have experienced a bombardment that was at least as intense (12). Thus, the likelihood of a large  $n$  for the Earth is high. A large  $n$  also suggests that  $E$  could be very modest, and yet the final  $P'_a$  might still approach unity.

Fig. 1 illustrates in simple terms the range of probability amplifications ( $E^n$ ) that result from exchange histories with  $n = 100, 1,000, 10,000$ . Here, the  $E$  factor ranges from close to unity to 1.1 (10% enhancement per exchange) and is chosen to illustrate the

transition between “modest” and extreme total amplifications. The horizontal line at  $E^n = 10^6$  is drawn to indicate the division between regimes where amplification can play a critical role in deciding between the plausible extremes for  $\lambda$ , as determined by Eq. 1; in other words, if the plausible range of abiogenesis probability (denoted by either  $\lambda$  or  $P_a$ ) based on the instance of the Earth were amplified by impact-exchange, what amplification factor represents the difference between no life on the Earth (at its present age) and life occurring within a few 100 Myr of formation?

It is clear that, for a scenario represented by Earth–Mars history, the likely exchange rate (of the order of 10,000 impact-driven opportunities) could readily place our solar system into a high-amplification category. We also note that if, for example, the growth of  $P_a$  instead followed a random walk (i.e.,  $E$  varied around unity in each exchange event), that would not preclude a net significant amplification. We therefore suggest that the natural exchange of chemical toolkits between just two young terrestrial-type planets could have an enormous influence on the overall likelihood of abiogenesis in a planetary system. Chemical exchange could in fact represent the difference between life arising or not. This mechanism could also provide a dominant source for variance in planetary abiogenesis probabilities, via differences in system architecture (e.g., orbital configurations of habitable planets and the population of potential impactors). An understanding of that variance might therefore be obtained through improved astronomical observations of exoplanet system properties.

If we assume that impact-exchange did indeed play a significant role in “amplifying”  $P_a$ , then Eqs. 1 and 2 should be reconsidered because they actually already incorporate the amplified probability implicitly. Thus, we can turn the argument around and propose that both this work and (for example) that of ref. 1 might only constrain the impact-exchange-enhanced  $P_a$  or  $\lambda$ . Thus, if  $10^{-3} \leq \lambda \leq 10^3 \text{ Gyr}^{-1}$  represents a plausible range for rare and not-rare life on suitable planets, and this probability has already been amplified by exchange the probability of life arising per unit time on a planet, the values of  $\lambda$  and  $P_a$  for planets *without* impact-exchange must be exceedingly small. For example, taking the above amplification of  $\sim 10^{30}$ , the probability of any life to have occurred ( $N_{\text{abiogenesis}} \geq 1$ ) during the entire history of a 4.5-Gyr-old Earth-analog planet without impact-exchange would be effectively zero (following Eq. 2).

It is also worth considering the implications for Martian abiogenesis. The amplification for the OoL on Earth would apply to early Mars too, but as of yet, we have no evidence for either extinct or extant lifeforms on Mars. Current analyses of remote and in situ data on Mars’s climate history indicate a strong probability of an extensive hydrosphere during the Noachian period (4.1–3.7 Gya), tapering off to the present cold and arid state by perhaps the Amazonian period ( $\sim 3$  Gya) (13). Because the Noachian corresponds to the period of maximal impact-exchange between Earth and Mars, it seems that a similar window of opportunity for the OoL could have existed on both planets. If high amplification occurred, it would therefore suggest that abiogenesis should have happened on Mars during the Noachian.

We further note that the accumulation of large datasets on exoplanetary systems (primarily via NASA’s Kepler mission) has led to confidence in the statistical statement that the architecture and contents of our solar system are somewhat atypical (14). Specifically, we lack super-Earth mass worlds with orbital periods of less than 100 d, but we do harbor longer period giant planets. It remains to be seen what the implication is for material exchange in young planetary systems with multiple habitable planets—although it is possible that our solar system was also unusually active in this regard.

## Conclusions

The life that we know of on Earth is united by a common biochemical alphabet. In the above discussion, the atoms or molecules referred to as building blocks are expected to become functional members of an “assembly” event. This functionality is implicit, for example, in the way that  $\langle n_o \rangle$  is used. However, we cannot yet rule out the possibility of an alternate, or “shadow,” life operating with a different alphabet (e.g., an inorganic one) (15). We also cannot not rule out the possibility that abiogenesis occurred on Earth because of interactions between mutually exclusive building blocks that together acted to dynamically increase the evolvable information content of the prebiosphere. In other words, it is possible that the building blocks of life today are the product of a combination of actions of earlier sets of building blocks that could not “cross-assemble” (shadow and nonshadow chemistries) but that could, together, generate the necessary complexity and selection for organic abiogenesis. Eq. 1 makes no assumptions about  $P_a$ ; the equation is simply the probability of assembly per unit time per set of building blocks when all other necessary conditions are met, but a further exploration of the potential influence of mutually exclusive building blocks could prove instructive.

In summary, we have sketched out a top-down perspective on OoL questions that links microscale factors to macroscale factors

through the formulation of a simple, but instructive, equation for the frequency of abiogenesis in a planetary environment. Given the current explosive progress in the discovery and characterization of exoplanets, the factors that we suggest play a role in OoL likelihoods could motivate the acquisition of specific observational data (e.g., the configuration and material exchange fluxes of planets per system, the early dynamical history of exoplanetary systems and differences from the solar system, and their chemical composition/heterogeneity). The approach encapsulated in our OoL frequency equation may therefore be useful for stimulating future discussions but should also have a genuinely quantitative role in studying the origins of life.

**ACKNOWLEDGMENTS.** The authors would like to thank Sara I. Walker (Arizona State University) for useful comments on the manuscript. This work stems from discussions held at the Earth Life Science Institute (ELSI) at the Tokyo Institute of Technology, as part of the ELSI Origins Network (EON) road-mapping workshop “A Strategy for Origins of Life Research.” C.S. thanks EON for financial support and hosting at ELSI, made possible by a grant from the John Templeton Foundation. C.S. also acknowledges the support of Columbia University’s Research Initiatives in Science and Engineering and NASA’s Nexus for Exoplanet System Science consortium via NASA Grant NNX15AK95G during the course of this work. L.C. thanks the Engineering and Physical Sciences Research Council (EPSRC) for financial support through Grants EP/L023652/1 and EP/J015156. L.C. also acknowledges the support of the University of Glasgow.

1. Spiegel DS, Turner EL (2012) Bayesian analysis of the astrobiological implications of life’s early emergence on Earth. *Proc Natl Acad Sci USA* 109(2):395–400.
2. Benner SA, Devine KG, Matveeva LN, Powell DH (2000) The missing organic molecules on Mars. *Proc Natl Acad Sci USA* 97(6):2425–2430.
3. Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth’s biogeochemical cycles. *Science* 320(5879):1034–1039.
4. Carter B (1983) The anthropic principle and its implications for biological evolution. *Philos Trans R Soc Lond A* 310:347–363.
5. Maccone C (2011) A mathematical model for evolution and SETI. *Orig Life Evol Biosph* 41(6):609–619.
6. Cirković MM (2004) The temporal aspect of the drake equation and SETI. *Astrobiology* 4(2):225–231.
7. Lineweaver CH, Davis TM (2002) Does the rapid appearance of life on Earth suggest that life is common in the universe? *Astrobiology* 2(3):293–304.
8. de la Oliva AR, et al. (2012) Assembly of a gigantic polyoxometalate cluster (W<sub>200</sub>Co<sub>80</sub>O<sub>660</sub>) in a networked reactor system. *Angew Chem Int Ed Engl* 51(51):12759–12762.
9. Gladman BJ, Burns JA, Duncan M, Lee P, Levison HF (1996) The exchange of impact ejecta between terrestrial planets. *Science* 271(5254):1387–1392.
10. Meyer C, et al. (2011) Shock experiments in support of the Lithopanspermia theory: The influence of host rock composition, temperature, and shock pressure on the survival rate of endolithic and epilithic microorganisms. *Meteorit Planet Sci* 46(5):701–718.
11. Ryder G, Koeberl C, Mojzsis SJ (2000) *Heavy Bombardment on the Earth at ~3.85 Ga: The Search for Petrographic and Geochemical Evidence. Origin of the Earth and Moon, 475 Tucson* (Univ. of Arizona Press, Tucson, AZ).
12. Burt DM, Knauth LP, Wohletz KH (2008) The late heavy bombardment: Possible influence on Mars, *LPI Contributions No. 1439* (Lunar and Planetary Institute, Houston), pp 23–24.
13. Clifford SM, Parker TJ (2001) The evolution of the Martian hydrosphere: Implications for the fate of a primordial ocean and the current state of the Northern Plains. *Icarus* 154:40–79.
14. Batygin K, Laughlin G (2015) Jupiter’s decisive role in the inner Solar System’s early evolution. *Proc Natl Acad Sci USA* 112(14):4214–4217.
15. Davies PCW, Lineweaver CH (2005) Finding a second sample of life on earth. *Astrobiology* 5(2):154–163.