

Metabolism, information, and viability in a simulated physically-plausible protocell

By: **Kristoffer R. Thomsen**^{1*}, **Artemy Kolchinsky**^{2,3†}, **Steen Rasmussen**^{1,4,5‡}

¹ University of Southern Denmark, 5230 Odense, Denmark

²ICREA-Complex Systems Lab, Universitat Pompeu Fabra, 08003 Barcelona, Spain

³ Universal Biology Institute, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

⁴ Santa Fe Institute, New Mexico 87501, USA

⁵ European Centre for Living Technology, 30123 Venice, Italy.

Abstract

Critical experimental design issues connecting energy transduction and inheritable information within a protocell are explored and elucidated. The protocell design utilizes a photo-driven energy transducer (a ruthenium complex) to turn resource molecules into building blocks, in a manner that is modulated by a combinatorial DNA-based co-factor. This co-factor molecule serves as part of an electron relay for the energy transduction mechanism, where the charge-transfer rates depend on the sequence that contains an oxo-guanine. The co-factor also acts as a store of inheritable information due to its ability to replicate non-enzymatically through template-directed ligation. Together, the energy transducer and the co-factor act as a metabolic catalyst that produces co-factor DNA building blocks as well as fatty acids (from picolinium ester and modified DNA oligomers), where the fatty acids self-assemble into vesicles on which exterior surface both the co-factor (DNA) and the energy transducer are anchored with hydrophobic tails. Here we use simulations to study how the co-factor sequence determines its ‘fitness’ as reflected by charge transfer and replication rates. To estimate the impact on the protocell, we compare these rates with previously measured metabolic rates from a similar system where the charge transfer is directly between the ruthenium complex and the oxo-guanine (without DNA replication and charge transfer). Replication and charge transfer turn out to have different and often opposing sequence requirements. Functional information of the co-factor molecules is used to probe the feasibility of randomly picking co-factor sequences from a limited population of co-factors molecules, where a good co-factor can enhance both metabolic biomass production and its own replication rate.

*krt.workmail@gmail.com

†artemyk@gmail.com

‡steensantafe@gmail.com

Contents

1	Introduction	2
2	Proto-cell functionalities and design	3
3	Simulation Exploration of Co-factor Charge Transport	9
3.1	Modelling charge transport reactions	11
3.2	Charge transport simulations in a selection of strands	12
3.3	Metabolic impact of DNA charge transfer	17
4	Simulation Exploration of Co-factor Replication	19
4.1	The LIDA system for protocells	20
4.2	Kinetic rates in LIDA	24
4.3	LIDA simulations for a selection of strands	27
5	Connecting results from charge transport and replication simulations	29
5.1	Functional characterization of co-factors	32
5.1.1	Direct rate combination	32
5.1.2	Functional information	33
6	Discussion	36
7	Conclusion	39

1 Introduction

In this work, we use simulations to explore simple molecular mechanisms that, when combined, can function as protocellular metabolisms and inheritable information systems. Both of these functionalities are critical to create ‘living’ materials from ‘nonliving’ materials. We may define a system to be ‘alive’ [1] if it can use free energy to metabolize resources into building blocks, such that the system can grow and divide to restart its ‘life-cycle’. Furthermore, if metabolism is regulated in part by inheritable information that can change between generations, then selection and simple evolution is possible. In this work, we define a protocell as a minimal physicochemical system that satisfies this operational definition, and we determine under which conditions such a protocell could be ‘viable’ with a combined metabolic-informational system.

Our bottom-up protocell approach builds on top of a diversity of related designs and methods that have steadily narrowed the gap between nonliving and living matter, e.g., see team efforts by Szostak et al. (2016) [2], Yomo et al. (2016) [3], Cronin et al. (2017) [4], Mann et al. (2017) [5,6], Rasmussen et al. (2016) [7], Adamala et al. (2022) [8], the Schwillie et al. and the MaxSynBio consortium (2018) [9,10], and perhaps the most ambitious, the ongoing Dutch Synthetic Cell project (2017) [11,12] headed by Dogterom. These bottom-up approaches are complementary to top-down approaches that modify existing living cells, as well as synthetic biology approaches that employ the biocatalytic machinery of contemporary life, e.g. see Venter team efforts (2006-now) [13].

Many experimental, theoretical and simulation studies in the literature have addressed challenges in the development of bottom-up protocells, including issues regarding container growth and division [14,15], container functionalities [8], formation of autocatalytic metabolic sets [16–18], information replication [19–21], thermodynamics of replication [22–26], protocell integration [7,19,27], inter-protocell interactions [5,6] as well as coupling and synchronization of growth among the different protocell processes [28,29]. A comparative discussion of many of these different approaches can be found in Rasmussen et al. (2008) [30].

In all bottom-up approaches, the greatest challenge is to integrate metabolism, information, and container into a functional and autonomous whole that can survive in some environment. Therefore, examining and understanding how novel functionalities emerge in complex physicochemical systems is crucial. This paper reports a continuation of ongoing work exploring how to assemble fully autonomous protocells in the lab with a metabolism design that contains an integrated energy transducer and a combinatorial co-factor [27,29,31–39].

In Section 2, we present and discuss the key features of the proposed protocell design, as well as the functionalities essential for the following analysis. In Sections 3 and 4, we use

simulations, which are motivated both by empirical observations and first-principles theory, to explore co-factor charge transfer and replication, which is the main focus of this paper. We conduct a simulation analysis of these two processes as they are not yet implemented in the laboratory; they are the two ‘missing links’ in a full laboratory implementation of our protocell design [36]. The simulations in Sections 3 and 4 may therefore be seen as preparations for future experiments. In Section 5, we compare and contrast the results from Sections 3 and 4, to rank the co-factor sequence efficiency for charge transfer and replication together with a combination of the two processes. Finally, we apply functional information [40,41] to our findings to get a sense of how likely it is to pick a ‘good’ co-factor at random from a limited DNA sequence ensemble. Section 6 critically discusses our findings, and in Section 7 we draw the conclusions from our study.

2 Protocell functionalities and design

The protocell is built around a metabolism that is directly coupled to an informational system and a container, all placed in an environment with appropriate access to resources and free energy. The design is “systemic” in the sense that the metabolic, information and container components are designed to mutually support each other [27]. The metabolic system utilizes light driven ruthenium tris(bipyridine) [Ru(bpy)₃] complexes (further referred to as Ru-C) as energy transducers and DNA as co-factors. Both the energy transducer and co-factor are tethered with hydrophobic anchors to the exterior surface of the fatty acid vesicles, which act as 2D containers; see Figure 1.

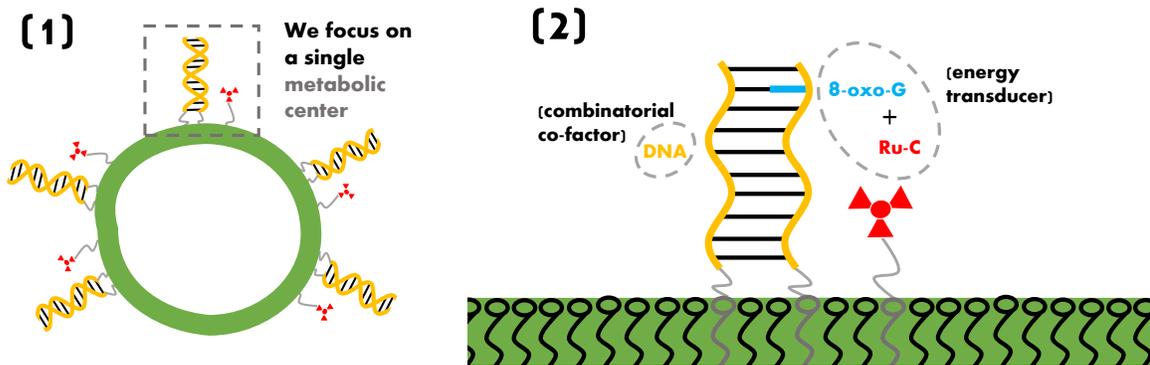


Figure 1: Panel (1) shows a full protocell. Panel (2) depicts the metabolic system for the protocell that is composed of a combined energy transducer (ruthenium complex) and a co-factor (DNA) that modulates the energy transduction process. The overall metabolic efficiency depends on the combinatorial properties of the DNA co-factor; different nucleobase composition and sequences alters the efficiency.

The combined energy transducer and cofactor, a two-component metabolic system, drives redox reactions on precursors of both amphiphiles (picolinium ester) and information molecules

(protected DNA oligomers). In this way, it produces self-assembling vesicles containing building blocks, decanoic acid, and functional co-factor DNA information templates. Thus, our protocell design is in several ways different from modern life: it uses non-biological energy transducers; it has both the energy transducers and the information (co-factor) molecules anchored on the outside of a vesicle (and not inside in the vesicle lumen); it does not have a DNA translation machinery; it uses no enzymes; and it uses simple fatty acids and not phospholipids for containers. A detailed illustration of the key steps in the protocellular processes are found in [Figure 2](#), [Figure 3](#), and [Figure 4](#).

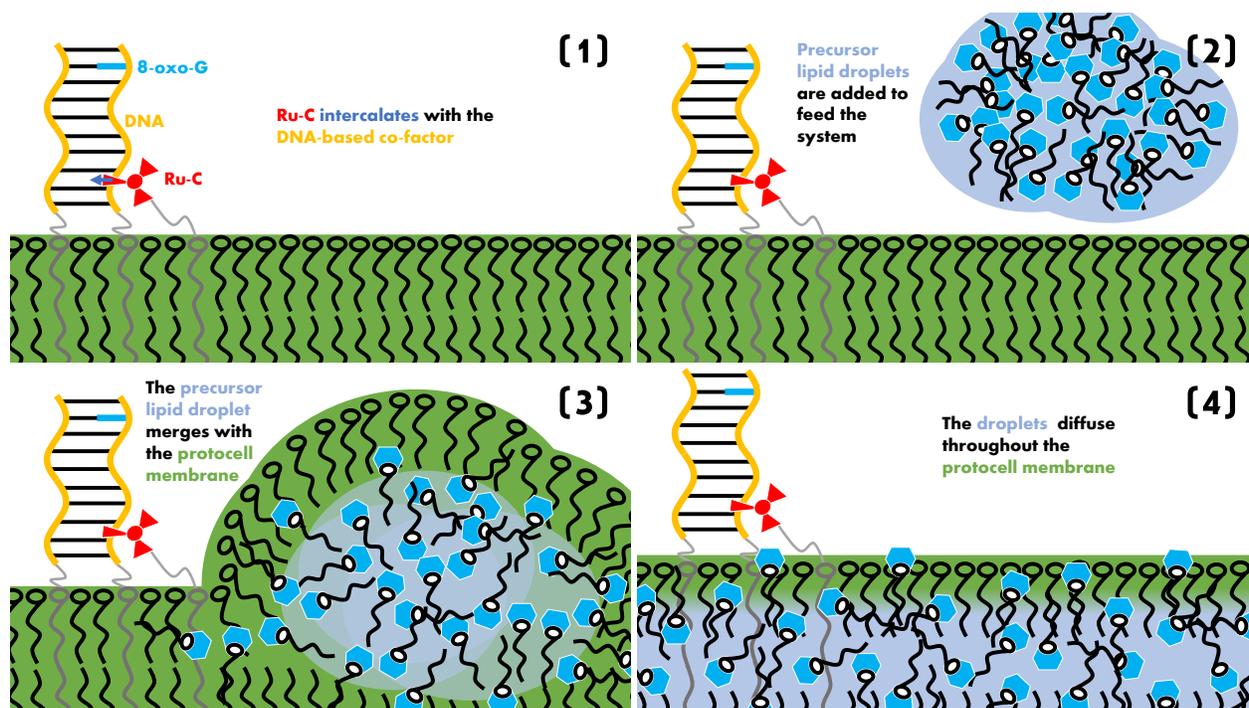


Figure 2: Membrane feeding: Panel (1) depicts the initial condition as discussed in [Figure 1](#), where the ruthenium complex is intercalated into the DNA duplex. Panel (2) shows that hydrophobic picolinium ester droplets are fed to the portocells via the environment. Due to their hydrophobisity, these droplets are incorporated into the fatty acid membrane where they initially tend to form globular membrane structures, see panel (3) [34]. Over time, some of the picolinium ester molecules tend to spread and organize within the membrane, Panel (4).

The two-component catalytic metabolism works as follows: A ruthenium-based complex (Ru-C) is activated by visible light and converts light energy (photons) into chemical energy (electrons) via two electron transfer processes (i) involving the DNA duplex with a modified nucleobase 8-oxoguanine (oxo-G) and (ii) the presence of a resource precursor molecule that has an easily breakable ester bond. (i) The activated and intercalated Ru-C receives an electron from a nearby guanine within the DNA duplex, which results in an electron hole in the π stacked duplex. This electron hole diffuses up and down the DNA π stack and eventually reaches the 8-oxo-G where it is absorbed due to the properties of the 8-oxo-G electron donor.

As a result of this electron donation, the 8-oxo-G loses a hydrogen atom, which is replenished by dihydrophenyl glycine as a hydrogen donor [33]. (ii) The light-activated electron that stems from the ruthenium atom is located at one of the three bipyridine molecules in the ruthenium complex. This electron may now jump to one of the picolinium ester molecules within the membrane and as a result deprotect its picolil group to form a new decanoic acid molecule, which serves as the building block of the fatty acid membrane. To ensure charge balance, a dihydrophenyl glycine is also needed as a hydrogen donor in this process [33].

A detailed analysis of the DNA charge (electron hole) transfer properties is the topic of Section 3.

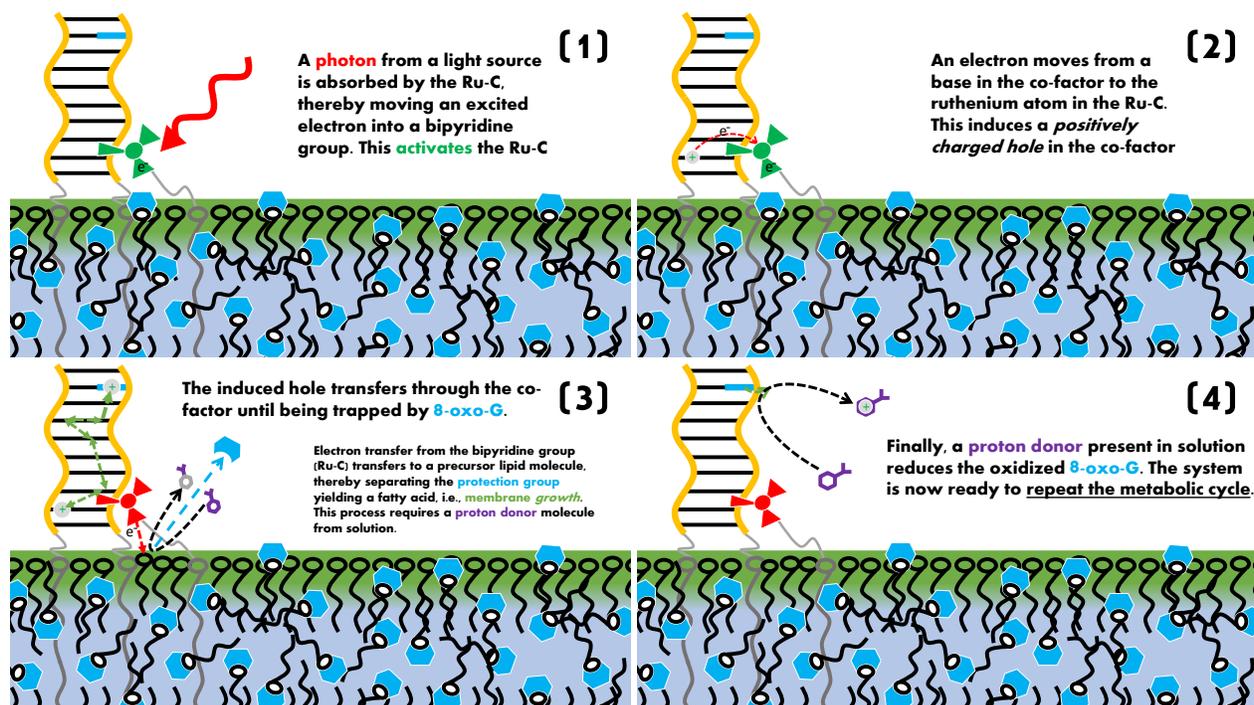


Figure 3: Metabolic membrane formation: Panel (1) depicts a photo-induced excitation of the ruthenium complex where an electron from the ruthenium atom jumps out onto one of the bipyridine rings. As a result, an electron from a nearby guanine molecule may now replace the missing ruthenium atom electron; see Panel (2). Panel (3) depicts (i) how the electron hole (the missing electron) within the DNA duplex can diffuse along the DNA duplex due to the π stacking of the base pairs. It should be noted that DNA hole transfer is sensitively dependent on the details of the DNA duplexes involved, which is the topic of Section 3 of this paper. Further, (ii) the originally excited electron may now jump from the bipyridine ring onto a nearby picolinium ester molecule, causing a fragmentation that results in a newly formed fatty acid and a free picolil group. Panel (4) depicts (i) trapping of the diffusing hole by the 8-oxo-guanine, resulting in the release of a hydrogen atom and replacement by a hydrogen donor (ii), so that the 8-oxo-guanine is restored and ready to receive another hole (donate another electron). For more details on the ruthenium chemistry involved, see [33] and [38]

DNA is used as a combinatorial co-factor in the metabolism because of its well known replication property. As the DNA base composition and sequence critically impact the metabolic efficiency, a protocellular inheritance is obtained by replication of the co-factor.

Due to the well-known product inhibition in template-directed ligation replication [42], we apply DNA sequences that are capable of self-replication via isothermal lesion-induced DNA amplification (LIDA) [43]. A detailed analysis of the co-factor replication properties is the topic for Section 4 of this paper.

Note that the oxo-G is located above the ligation site to ensure that charge transfer can occur only in conjunction with DNA replication that requires hybridization and ligation into a duplex strand. In that way, the replication process becomes directly linked to the metabolic process. On the contrary, if the oxo-G is located near the vesicle surface at the lower end of the DNA strand, direct charge transfer may become possible between the ruthenium complex and the oxo-G without necessitating DNA hybridization and ligation. Thus, the metabolism would depend only on the presence of an anchored precursor oligomer containing oxo-G. For simplicity, in the following, we require the oxo-G to be located above the ligation site. However, a weaker linkage between replication and metabolism might still be possible with an oxo-G located below (but close to) the ligation site, as charge transfer would also likely require hybridization (but not ligation) and thus an initiation of the replication process.

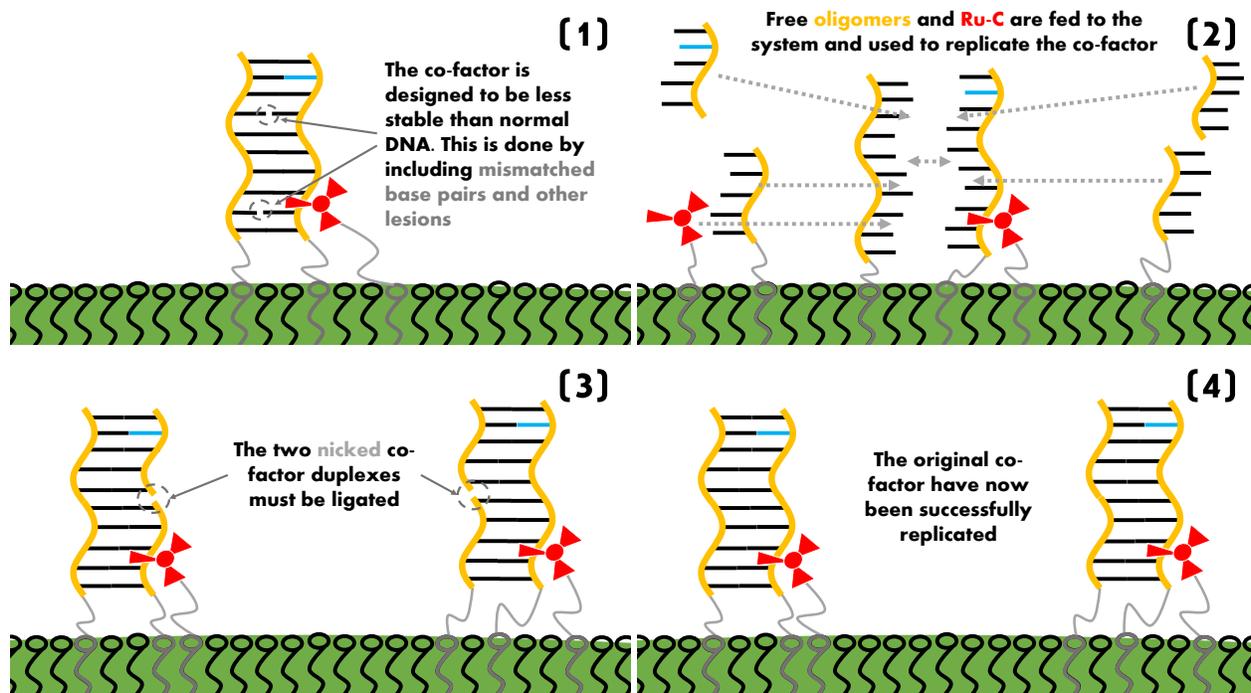


Figure 4: Co-factor replication: Panel (1) depicts the same initial conditions as shown in [Figure 2](#) and [Figure 3](#), although an extra detail is added as the involved DNA strands must contain lesions such as mismatches, abasic sites or bulges to destabilize the duplex. Panel (2) depicts (i) feeding of four oligomers and ruthenium complexes from the environment. The ruthenium complex and two of the oligomers have anchors. (ii) Due to the dynamic equilibrium between hybridized and dehybridized oligomers and the full strands, templating between the original strand, its complementary strand, and the newly added oligomers is possible. It should be noted that only DNA sequences with particular properties (destabilizing mismatches, abasic sites, or bulges) can replicate in this manner, which is the topic for Section 4 of this paper. In Panel (3), a full hybridization between the oligomers and the templates has occurred, where ligation can occur directly as the 3' end is activated by imidazole in the situation where no oligomer protection group is present at the 5' end. A hypothetical oligomer protection group could be removed if the protection group bond were disrupted by the traveling hole (positive charge) inside the DNA duplex. Oligomer deprotection is previously demonstrated by an excess electron (not a hole) [35] where the activated electron is also provided by ruthenium complex photo-activation. Panel (4) depicts the resulting replication of the co-factor.

To summarize: The combinatorial DNA co-factor system serves as part of an electron relay that modulates the energy transduction efficiency, which directly depends on the charge transfer rates of the co-factor DNA information molecules. Further, non-enzymatic replication of the combinatorial DNA co-factor and added ruthenium complexes ensure continued metabolic functionality in the next generation. Note that both DNA charge transfer and non-enzymatic DNA replication strongly depends on the DNA sequence composition. As we shall see in the following Section, the DNA charge transfer and replication properties have very different sequence requirements, which makes it nontrivial to identify appropriate sequences that can support both functionalities.

Further, it should be noted that the environment is highly regulated (e.g., temperature, pH, salts) and designed to support the protocellular life-cycle by delivering light energy as

well as multiple necessary resource molecules including picolinium ester, ssDNA oligomers (with and without tails), Ru-complexes with tails, picolinium ester, dihydrophenyl glycine (hydrogen source). A summary discussion of the main processes in the protocellular life-cycle is shown in Figure 5

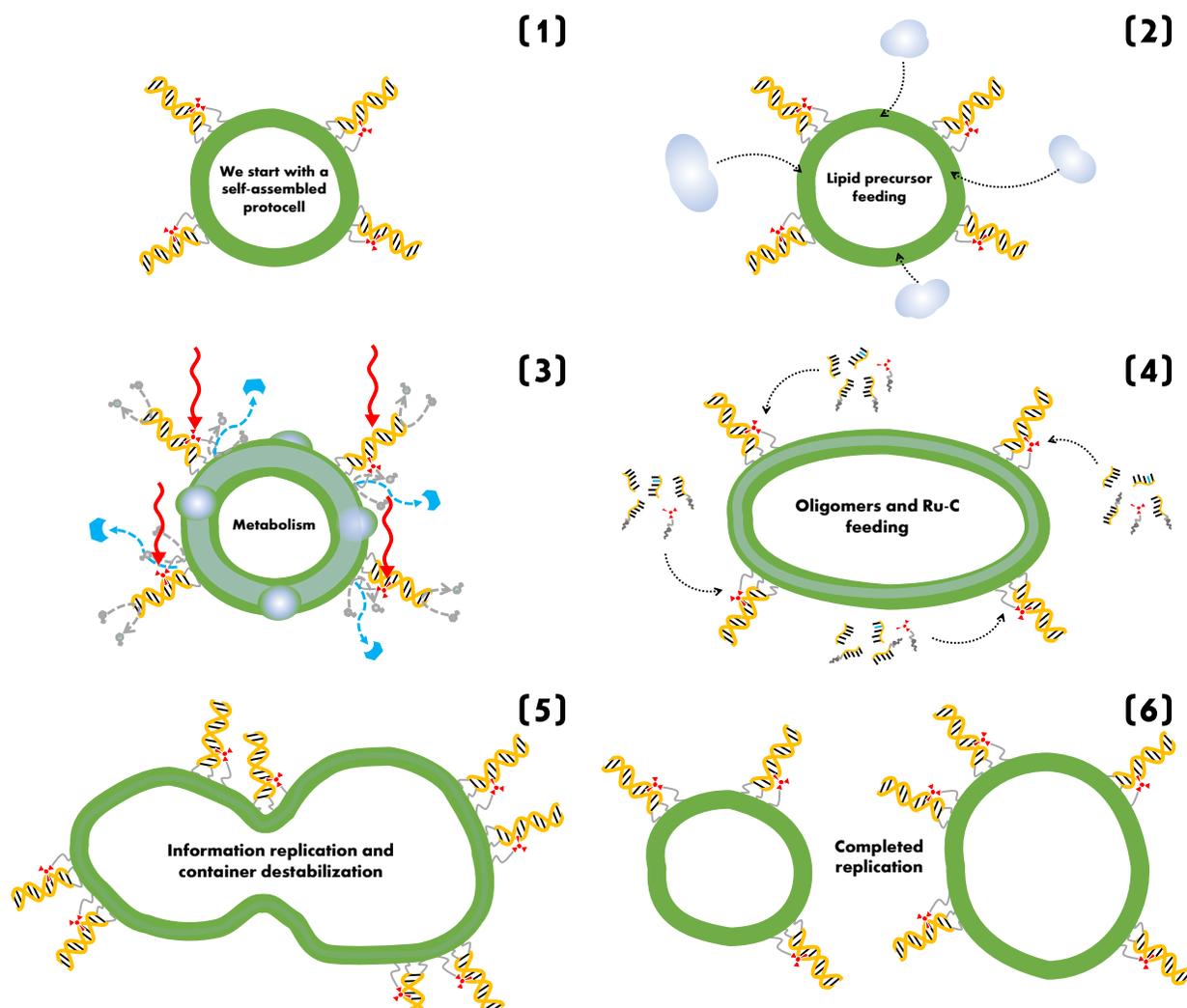


Figure 5: Protocellular life-cycle: Panel (1) is a cartoon of the protocell with a fatty (decanoic) acid vesicle container decorated with anchored ruthenium complexes and DNA duplexes. Panel (2) shows droplet feeding of hydrophobic membrane precursors (picolineum esters). Panel (3) depicts absorption of the hydrophobic droplets into the membrane, where they are partly dissolved in the membrane, while the metabolism converts precursor lipids (picolinium ester) into lipids (decanoic acid). Panel (4) depicts ruthenium complex and the feeding of the DNA oligomer. Note that all precursors can be fed at once but are shown here as a two-step process for clarity. Panel (5) shows a completed (DNA) co-factor replication together with a completed fatty acid production, which results in membrane growth and vesicle destabilization. The original vesicle eventually breaks up and forms two new protocells, see Panel (6).

3 Simulation Exploration of Co-factor Charge Transport

Charge transport (CT) in DNA has been extensively studied since the early 1990s. The majority of DNA CT research has focused on hole transport (HT). Less research has been done on electron transport (ET) in DNA [44]. DNA HT dynamics is usually associated with oxidized nucleobases in DNA. Oxidative damage occurs through the oxidization of guanines by reactive oxygen species leading to higher levels of 8-oxo-7,8-dihydro-2'-deoxyguanosine (oxoG), which is linked to mutagenesis and cancer [45, 46]. Therefore, electrochemistry-based sensors have been proposed for the detection of selected DNA sequences or mutated genes associated with human disease [47].

Our previous explorations of protocellular metabolic function have focused on direct ET from oxoG to a Ru-C, either covalently bound or in close spacial proximity [33–36, 38]). In the present work, the direct ET between oxoG and Ru-C is mediated by a combinatorial co-factor in the form of a DNA duplex. This causes CT to become a multi-step process, that both involves ET and HT. Electron transport from oxoG through DNA to Ru-C is equivalent to: (i) An initial ET from a nearby guanine within the DNA strand to the light activated Ru-C; (ii) Now the missing electron in the guanine (the hole) travels through the DNA strand HT until it meets an oxoG that donates the missing electron. To repair the now damaged oxoG, a proton donor (dihydrophenylglycine) is provided to the system as in our previous studies.

Everything else being equal, the slower the CT rate within the DNA, the slower the resulting metabolic rate, because the CT rate modulates the rate at which photo-excited electrons can be used for digesting resource molecules and turning them into building blocks.

Previous studies have established that Ru-C can merge into the π stack of DNA duplexes (intercalation), which can induce HT into DNA when photo-activated. It is known that Ru-C intercalates with the π stack of DNA duplexes, yielding fast photoinduced transfer of a hole to DNA nucleobases, and such holes can migrate long distances (200 Å) through DNA [48–50]. It is also known that HT occurs in DNA strands containing oxoG, and that oxoG is a deep thermodynamic potential well for holes [51]. Thus, holes become ‘trapped’ when reaching oxoG.

A design example of the protocellular metabolism, composed of the energy transducer and the co-factor, is shown in **Figure 6**. Based on these experimental facts, we propose a protocell design in which DNA strands containing a single oxoG near one terminal is attached to the protocell container (membrane) via at least one amphiphilic anchor at the opposite terminal. Additionally, Ru-C (energy transducers) are anchored to the membrane

via molecular tethers, thereby promoting intercalation with the π stack of the DNA duplex.

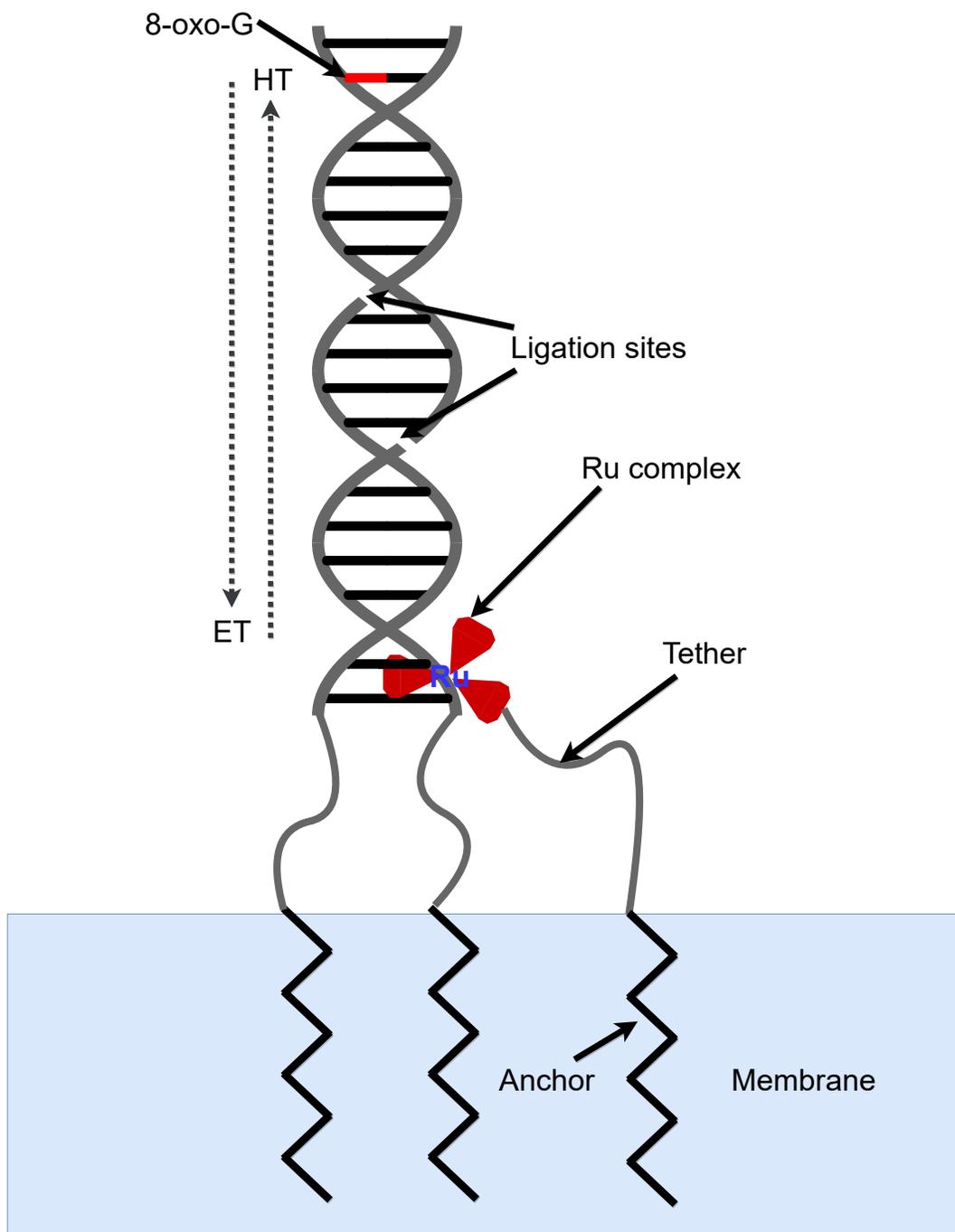


Figure 6: Example of protocellular metabolism design. The co-factor containing oxoG and Ru-C are anchored to the protocellular container surface via tethers to amphiphilic anchors. The Ru-C intercalates with the DNA π stack of the co-factor enabling CT between Ru-C and 8-oxo-G mediated by the nucleobases of the DNA strand, see text for details. A fast DNA HT means a fast metabolism, while a slow HT means a slow metabolism. Examples of DNA oligomer ligation sites are shown as they impacts the HT rates.

3.1 Modelling charge transport reactions

Because the efficiency of the DNA CT is critical for metabolic efficiency, we need to explore how the CT is influenced by the DNA base composition. We can model DNA CT as a system of multiple redox centers, which can be modeled as systems of coupled linear differential equations. A system with N redox centers is described by N equations of the following form:

$$\frac{dP_i(t)}{dt} = \sum_{j=1}^N k_{ij} P_j(t) \quad \text{where } i \in \{1, \dots, N\}, \quad (1)$$

where $P_i(t)$ describes the probability that the charge is located at the i th redox center at time t and k_{ij} are elements of a matrix containing the transition rates between redox centers: k_{ij} being the transfer rate *to* the i th *from* the j th redox center. Assuming there is no loss of charge to the environment, the rate matrix should conserve charge, meaning $\sum_i k_{ij} = 0$.

We have chosen to restrict our kinetic simulations in the following sections to use previously calculated or experimentally measured CT rates.

We now use the rate matrix formalism to model CT in DNA. Because the transition rates depend on quantum tunneling rates, they decay exponentially as a function of the distance between redox centers. Models that only include thermally induced hole transport transitions (TIH) set entries far from the diagonal of the k_{ij} matrix to zero, yielding a band matrix. If transitions to non-neighboring bases are allowed — e.g., by including the super exchange (SE) mechanism — the rate matrix will no longer be a simple band matrix, but instead have non-zero entries away from the band, as is the case for the rate matrix in Eq. 2 that represents a simple well-matched strand 5'-GAGAAGA-3':

$$\mathbf{K} = \begin{bmatrix} -k_{GA} & k_{AG} & 0 & 0 & 0 & 0 \\ k_{GA} & -(2k_{AG} + k_{SE}) & k_{GA} & 0 & k_{SE} & 0 \\ 0 & k_{AG} & -(k_{GA} + k_{AA}) & k_{AA} & 0 & 0 \\ 0 & 0 & k_{AA} & -(k_{AA} + k_{GA}) & k_{AG} & 0 \\ 0 & k_{SE} & 0 & k_{GA} & -(2k_{AG} + k_{SE}) & k_{GA} \\ 0 & 0 & 0 & 0 & k_{AG} & -k_{GA} \end{bmatrix} \quad (2)$$

For the construction of these models, the following assumptions are made: (i) holes are localized at a single base; (ii) HT transitions are always possible between neighboring bases; (iii) holes only transfer between the bases with the lowest oxidation potential in each base pair ($E_{ox}^G < E_{ox}^C$ and $E_{ox}^A < E_{ox}^T$ [52]); (iv) all SE transitions — independent of bridge sequence — are modelled as SE over a simple A/T bridge, i.e., the SE rate from G to GGG

found by Bixon and Jortner is approximately equal to the SE rate from G to a single G¹; (v) there is no charge loss from the strand to solution. Furthermore, in our simulation, we set a maximum bridge length for SE transitions of $n = 9$ bp due to the rapid exponential decay of SE rates as a function of n . For all future simulations the charge is initially localized at the first base, i.e., $P_1(t = 0) = 1$ and $P_{i \neq 1}(t = 0) = 0$.

3.2 Charge transport simulations in a selection of strands

To obtain physically realistic co-factor CT rates we only simulate strands with known experimental or theoretical single-step HT k_{ij} rates. These experimentally estimated rates are from Osakada et al. (2008) [53], while the theoretically estimated rates are from Bixon and Jortner (2002) [52]. The used theoretical rates computed by Bixon and Jortner, who use semi-classical Marcus theory are shown in Table 1. In this table are also shown the rates from Osakada et al. (2008) that experimentally investigate charge transport through strands containing mismatches. From the charge transport they further deduce rates between certain redox sites (not necessarily neighboring base pairs) by fitting kinetic models to their results. As will become clear in Section 4.2 on co-factor replication, we have selected a group of strands containing both a bulge (an unpaired “bulging” base) and possibly mismatches (bases forming non-Watson–Crick pairs) [54]. A bulge occurs when one base is not is detached from the duplex structure somewhere in the middle of the strand, e.g., if one strand of in a well-matched duplex is one base longer than the other strand.

Specifically, in our investigations, we assume the rate through the substrand containing the mismatch is independent of the sequence of the rest of the strand, so that rates across mismatches (k_{GGt} and k_{GGa}) are implemented as a combined rate across multiple base pairs. Furthermore, the charge transport across the mismatch is modeled as being one-directional (no back-transfer) as by Osakada et al. (2008). The values of k_{GGt} and k_{GGa} are also shown in Table 1.

Research by Barton and coworkers has found that conformational gating is highly regulating in DNA HT, with disruptions of the DNA π stack generally yielding slower HT rates [49,55]. This has motivated our estimate of HT rates through the bulge (k_{AB} and k_{BA}) as being approximately one order of magnitude slower than HT rates through mismatches, as we view mismatches as a lesser disruption of the DNA π stack compared to bulges.

¹This assumption is based on the energy dependent nuclear Franck-Condon factor in Eq. (31), where Bixon and Jortner [52] use $\lambda = 0.25$ eV and $\hbar\omega = 0.18$ eV, both of which are large compared to $\Delta G = -0.096$ eV.

Name	Rate [s ⁻¹]	Reference	Description
k_{AA}	$5 \cdot 10^7$	Bixon and Jortner (2002) [52]	rate from A/T to A/T
k_{GA}	$4 \cdot 10^4$	Bixon and Jortner (2002) [52]	rate from G/C to A/T
k_{AG}	$3 \cdot 10^8$	Bixon and Jortner (2002) [52]	rate from A/T to G/C
$k_{GG}(n = 1)$	$1.2 \cdot 10^{8a}$	Bixon and Jortner (2002) [52]	super-exchange rate from G/C to G/C
k_{AoxoG}	$4 \cdot 10^6$	Estimation by authors	rate from A/T to oxoG/C
k_{oxoGA}	0 ^b	Estimation by authors	rate from oxoG/C to A/T
k_{GGt}	$1 \cdot 10^4$	Osakada et al. (2008) [53]	rate through substrand from G/C to G/T
k_{GGa}	$9.5 \cdot 10^3$	Osakada et al. (2008) [53]	rate through substrand from G/C to G/A
k_{AB}	$1 \cdot 10^3$	Estimation by authors	rate from A/T to bulge
k_{BA}	$1 \cdot 10^3$	Estimation by authors	rate from bulge to A/T

Table 1: ^aThe super-exchange rate between G/C and G/C through an A/T-bridge of length $n = 1$. The super-exchange rates for A/T-bridges of length $n > 1$ are defined as $k_{GG}(n) = k_{GG}(1)r^{n-1}$, where r is determined by the electronic couplings in the bridge or alternatively a β -value specific for the bridge ($r = \exp(-\beta R_0)$). In our investigations, we use $\beta = 0.7 \text{ \AA}^{-1}$ and $R_0 = 3.4 \text{ \AA}$; R_0 being the base separation in the DNA π stack. In our simulations, we set a maximum bridge length of $n = 9$ bp. ^bSince oxoG is a thermodynamic trap, the rate out of oxoG is assumed to be very low compared to the rest of the used rates. k_{oxoGA} is therefore set equal to zero for simplicity.

Clearly, using only the elemental HT rates listed in Table 1 limits the possible strands we can simulate in a sequence of 18 bps. Eqs. (3)-(8) below show the six used substrands from which the full 18 bp co-factor sequences are composed. From Table 1, we have that (i) substrand 3 and 4 (mismatches) must be preceded by substrand 1 for the rate constants k_{GGa} and k_{GGt} to apply. Furthermore, (ii) the full strand must include one and only one bulge (substrand 6) near the center of the strand (see Section 4 for details). We choose to place it as the third substrand in our chosen set. Finally, (iii) the full strand must include one and only one 8-oxo-G (substrand 5). This results in a set of 160 strand sequences that are chosen as possible protocell co-factor candidates.

$$\text{strand 1: } 5'\text{-AGA-3'} \tag{3}$$

$$3'\text{-TCT-5'}$$

$$\text{strand 2: } 5'\text{-AAA-3'} \tag{4}$$

$$3'\text{-TTT-5'}$$

$$\text{strand 3: } 5'\text{-AGA-3'} \tag{5}$$

$$3'\text{-TTT-5'}$$

$$\text{strand 4: } 5'\text{-AGA-3'} \tag{6}$$

$$3'\text{-TAT-5'}$$

$$\text{strand 5: } 5'\text{-A8oxoGA-3'} \tag{7}$$

$$3'\text{-TCT-5'}$$

$$\text{strand 6: } 5'\text{-AAA-3'} \tag{8}$$

$$3'\text{-TT-5'}$$

The full 160 strands are shown in [Table 3](#) in [Section 5](#). In addition, in that table, the conclusions from [Sections 3](#) and [4](#) are indicated by color coding of the different strands.

For the protocell metabolism, the quantity of interest is the HT rate from Ru-C to oxoG (henceforth denoted k_{CT}). Note that this CT is coupled to the larger, full protocellular metabolism [[33–36, 38](#)], which is further discussed in [Section 3.3](#). For the following CT simulations, it is assumed that a Ru-C intercalates at the 5' end of the DNA strand, inducing a hole by fast HT into the first base of the 5' end. In principle, the hole could be induced at other sites instead. The hole then travels within the strand until it reaches oxoG, where it is trapped. A MATLAB program constructs the system of ODEs specific for each of the 160 strand sequences using the rate matrix formalism in [Eq. \(1\)](#) and solves the resulting system numerically with `ode15s`. As previously stated, the hole is assumed to be initially localized at the first base, the initial conditions $P_1(t = 0) = 1$ and $P_{i \neq 1}(t = 0) = 0$ are used.

One way of quantifying transport in diffusion-to-target reactions is the Median First Passage Time (MedFPT), which describes the median time it takes for a diffusing particle to reach a target position for the first time. We will now use MedFPT to quantify the rate of HT through the co-factor to 8-oxo-G. Since the model assumes 8-oxo-G to be a complete trap for holes (no transitions out of 8-oxo-G), the mathematics of identifying the MedFPT simplifies to identifying the time at which $P_{8oxoG}(t) = 0.5$, i.e., the time at which the probability of the hole being located at 8-oxo-G is 0.5. The inverse quantity $k_{CT} = 1/\text{MedFPT}$

then defines the rate of HT to 8-oxo-G. To illustrate the typical emergent hole dynamics, **Figure 7** shows the obtained temporal dynamics of CT for one of the simulated strands.

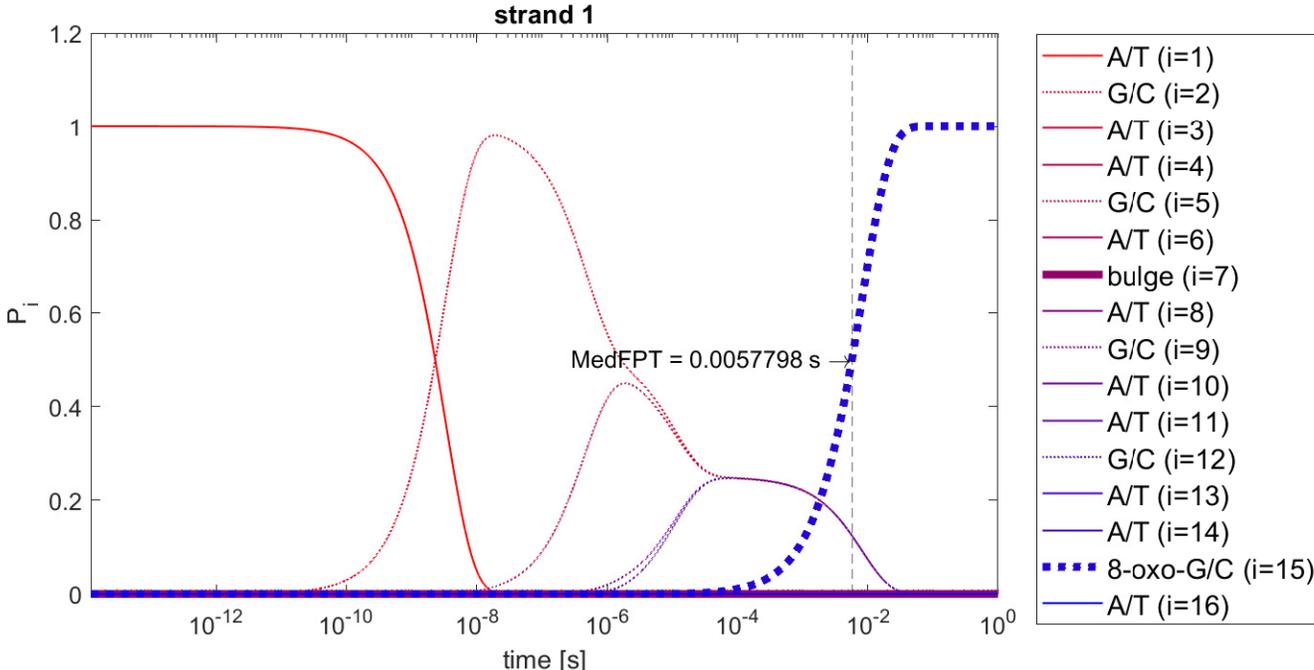


Figure 7: CT simulation results for strand 1 to determine the median first passage time (MedFPT). The dynamics show the influence of the fast SE transition between $G(i=5)$ to $G(i=9)$ which circumvents the slow transition through the bulge.

The following tendencies emerge from the 160 strand simulations: Primarily, the fastest CT rates are observed when oxoG is located before the bulge, as expected. Second, SE and its distance dependence highly influence CT rates, especially when SE occurs over the bulge. When no SE transitions across the bulge are present (when NN guanines are separated by more than 9 bp), the slowest total CT rates are observed. Mismatches are less impeding than expected for CT rates in the current implementation, possibly due to mismatch CT rates being implemented as a single transition over multiple bases, which is comparable to the rate of multiple thermally induced hole (TIH) jumps across the well-matched strand. Finally, the assumption of no back-transfer across mismatches yields cases where the stopping condition ($P_{8\text{oxoG}}(t) = 0.99$) is not fulfilled within the allotted simulation time, as charge becomes trapped away from oxoG. This assumption may also lead to mismatches increasing CT rates in cases where oxoG is located after mismatches, as charge is trapped by mismatches at close proximity to oxoG. **Figure 8** shows a histogram of all obtained CT rates for the 160 strands.

We also calculated the Mean First Passage Time (MeanFPT) for the same charge transport matrices. To calculate the MeanFPTs, we defined $\tilde{\mathbf{K}}$ as the rate matrix in Eq. 2 except with the row and column corresponding to the absorbing 8-oxo-G site removed.

The MeanFPT starting from initial site 1 is then given in terms of the matrix inverse as $\tau_{\text{MeanFPT}} = -\sum_i [\tilde{\mathbf{K}}^{-1}]_{i1}$ [56]. Generally, we find that $\text{MedFPT} \simeq 0.7 \times \text{MeanFPT}$, except for Strands 71 and 73 where $\text{MedFPT} \simeq 0.04 \times \text{MeanFPT}$. This difference is due to the special structure of these strands, as they both have the oxoG below the ligation site combined with a small transition probability for the hole to jump past the ligation site. Most holes move directly to the oxoG and get absorbed, while a few jump to the upper part of the strand where they get ‘stuck’ diffusing along the upper part of the strand before eventually moving past the ligation site (bulge) and getting absorbed by the oxoG.

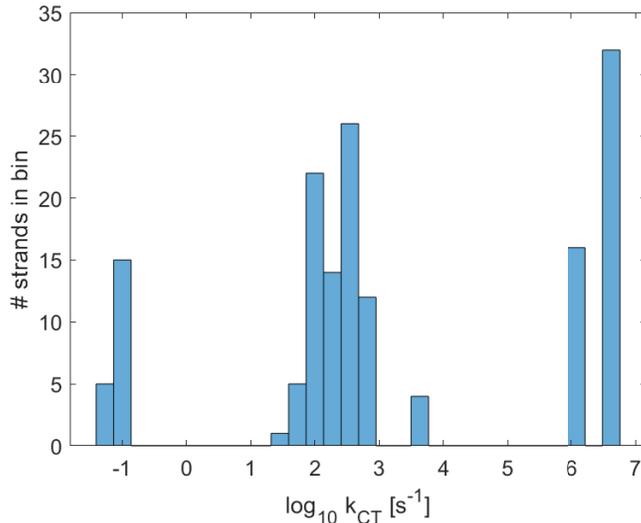


Figure 8: Histogram showing the distribution of charge transfer rates to 8-oxo-G through the set of 160 strands. Note the three grouping of the CT rates and see text for details.

The CT rate data k_{CT} shown in **Figure 8** show three distinct groups: (i) strands with $k_{CT} < 10^{-1} \text{ s}^{-1}$, strands with $10^0 < k_{CT} < 10^4$, and (iii) strands with $k_{CT} > 10^5 \text{ s}^{-1}$. The common factor for group (i) is the absence of a SE transition across the center bulge, which yields comparatively slower CT rates. Group (iii) is further divided into two subgroups corresponding to the 8-oxo-G being located in either the first or second substrand of the total strand, and the high observed CT rates are therefore simply due to the hole being initially located close to the 8-oxo-G.

As discussed in Section 2, an 8-oxo-G location before the ligation site is not preferable as the metabolic processes in this case can proceed with the supply of only the lower-resource oligmer. Without a full replication process, the metabolism thereby loses its information control.

Variations in group (ii) of strands with $1 \text{ s}^{-1} < k_{CT} < 10^4 \text{ s}^{-1}$ are due to an assortment of factors such as number of mismatches, mismatch locations, and distance dependence of SE

transitions. A more complete correlation analysis remains a future task.

3.3 Metabolic impact of DNA charge transfer

As discussed in Section 2, DNA charge transfer is explored as a sequence-dependent process that could be part of the network of protocellular metabolic processes. The key issue for our investigations is to understand how the DNA charge transfer process influences the overall metabolic performance.

DNA charge transfer impacts the rate by which the photo-activated electron can be utilized to transform resource molecules into building blocks. Therefore, for our purposes, it suffices to review the initial reactions in the metabolic reaction network, which are summarized in the upper left part of **Figure 9** above the blue dotted line. A detailed discussion of the full metabolic network can be found in [33,38,57], although this earlier work did not include the double stranded DNA charge transfer process (missing the step that includes the DNA⁺oxoG complex).

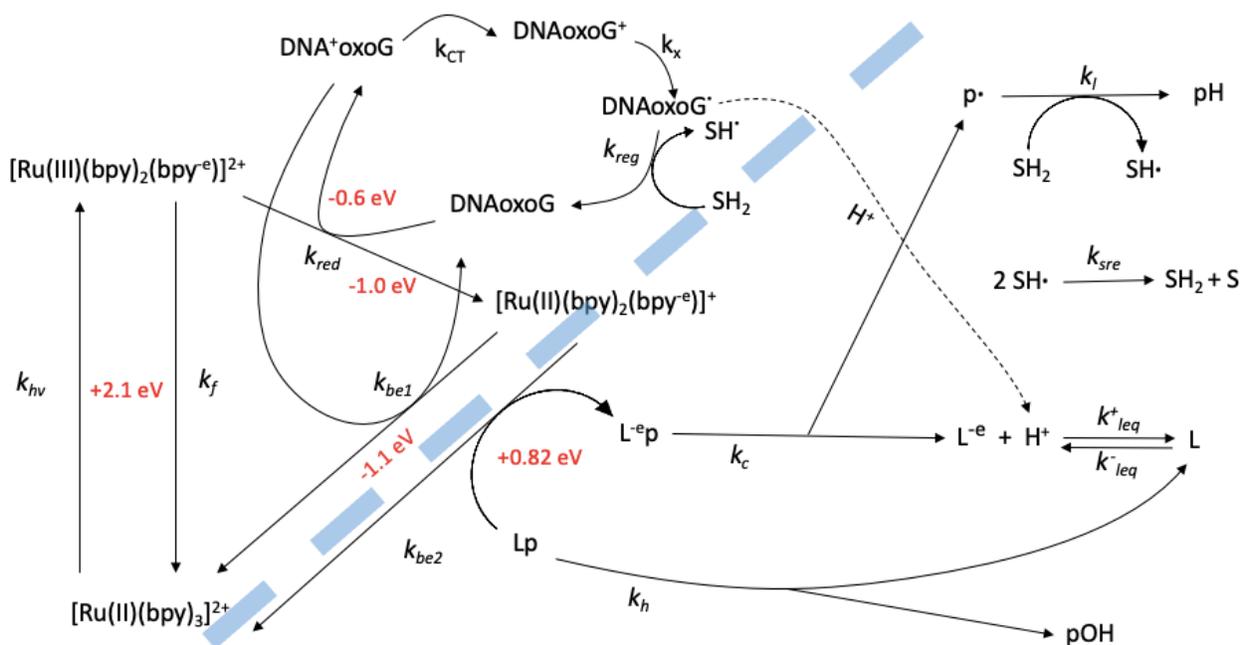


Figure 9: Protocellular metabolic reaction network. See text for details. Note that we only need to consider the upper left part (above blue dotted line) of the reactions to evaluate the overall impact of the DNA charge transfer process. Presentation and discussion of the full network and its kinetics can be found in [33,38,57].

In **Figure 9**, the reaction constant $k_{hv} \simeq 0.5/\text{s}$ determines the photo activation rate. The fluorescent rate constant is $k_f \simeq 3 \times 10^6/\text{s}$, while the bimolecular reduction rate constant is $k_{red} \simeq 3 \times 10^5/\text{M s}$. These rate constants are assumed to be constant in the following. Furthermore, if we assume that the back reaction for an electron already transferred from

the DNA helix to the ruthenium complex can be ignored (no covalent binding between the two reactants), we can set $k_{be1} \simeq 0$.

The circular reaction in the upper part of **Figure 9** starts with the DNA duplex donating an electron to the ruthenium complex to become DNA⁺oxoG. The rate of this reaction is determined by the bimolecular rate constant $k_{red} \simeq 3 \times 10^5/\text{Ms}$ and the ruthenium complex concentration $[\text{RuC}]$. Next, a hole diffuses inside the DNA⁺oxoG duplex toward the oxoguanine, whose rate constant k_{CT} we estimated for a variety of strand combination in this Section and summarized in **Figure 9**. The hole is eventually absorbed by the oxoguanine (DNAoxoG⁺), which thereby loses a proton, as determined by the rate constant $k_x \simeq 10^8/\text{s}$ to become DNAoxoG[•]. The oxoguanine, now missing a proton, is successively provided with a new proton from a sacrificial proton donor, SH₂ as dihydrophenylglycine, whose rate is determined by the bimolecular rate constant $k_{reg} \simeq 10^6/\text{M s}$ and the concentration $[\text{SH}_2]$. This concludes the cycle and makes the DNA duplex ready to donate a new electron to the ruthenium complex.

The reaction time $t_{circular}$ for complete this circular reaction is determined by the four involved reaction constants $k_{red}[\text{RuC}]$, k_{CT} , k_x , $k_{reg}[\text{SH}_2]$ and can be approximated by $t_{circular} \simeq 1/k_{red}[\text{RuC}] + 1/k_{CT} + 1/k_x + 1/k_{reg}[\text{SH}_2]$, so the resulting reaction rate constant can be approximated as

$$k_{circular} \simeq \{k_{red}[\text{RuC}] \times k_{CT} \times k_x \times k_{reg}[\text{SH}_2]\} / \left\{ (k_{CT} \times k_x \times k_{reg}[\text{SH}_2]) + (k_{red}[\text{RuC}] \times k_x \times k_{reg}[\text{SH}_2]) + (k_{red}[\text{RuC}] \times k_{CT} \times k_{reg}[\text{SH}_2]) + (k_{red}[\text{RuC}] \times k_{CT} \times k_x) \right\}. \quad (9)$$

If we assume that $k_{CT} \ll k_{red}[\text{RuC}]$, k_x , $k_{reg}[\text{SH}_2]$, we have $k_{circular} \simeq k_{CT}$. In this situation k_{CT} dominates the ‘recharge’ time of the DNA co-factor, determining how fast it can again act as an electron donor for the ruthenium complex. From the experiments in [33] we have $[\text{RuC}] = 1 \text{ mM}$ and $[\text{SH}_2] = 15 \text{ mM}$, which means that $k_{red}[\text{RuC}] \simeq 3 \times 10^5/\text{Ms} \times 10^{-3}\text{M} \simeq 300/\text{s}$ and $k_{reg}[\text{SH}_2] \simeq 10^6/\text{s} \times 15 \times 10^{-3}\text{M} \simeq 1.5 \times 10^3/\text{s}$. Thus, $k_{CT}(= 50/\text{s}) \ll k_{red}[\text{RuC}](= 300/\text{s})$ is only weakly fulfilled with these concentrations, as the ruthenium complex concentration is a secondary rate limiting factor for $k_{circular}$.

In any event, increasingly smaller k_{CT} means an increasingly smaller rate of usable photo activated electrons determined by

$$k_{usable-photo} \simeq \frac{k_{h\nu} \times k_{CT}}{k_{h\nu} + k_{CT}} \quad (10)$$

This means that depending on whether k_{CT} is the same as $k_{h\nu}$, or is 10 times larger, or is

100 times larger than k_{hv} , the amount of usable photo activated electrons will be reduced by about 50%, 10% and 1% respectively. *In the following, we require that $k_{CT} > 100 \times k_{hv} = 50/s$, so that the photo activation process is still the overall rate-limiting reaction.*

Going back to **Figure 8**, recall that the strands in group (ii) with the 8-oxo-G located above the ligation site are acceptable protocellular co-factor candidates. Furthermore, requiring that $k_{CT} > 50/s$ should ensure co-factors that make viable protocells.

The key question is; how fast, or slow, is the resulting fatty acid production rate if we accept the above requirements? From previous experimental studies [33,34], we measured an initial resulting fatty acid reaction rate constant of about $1.3 \times 10^{-5}/s$, which depends on the aggregate surface composition of picolinium ester (resource molecules) and fatty acid (products). More fatty acid membranes lower the reaction rate constant, presumably because the picolinium ester absorbed and integrated into the membrane is less accessible for the metabolic complex. There is, however, one more issue to consider, since a direct hydrolysis of the picolinium ester is also happening, which means that the production of fatty acid can also occur without the protocellular controlled photo-driven conversion. The hydrolysis rate constant for picolinium ester hydrolysis is measured up to about $10^{-6}/s$ [33], about an order of magnitude lower than the catalyzed charge transport driven rate constant of about $1.3 \times 10^{-5}/s$, recall the discussion above.

Assuming a rate constant of half the above and a 1:1 composition of picolinium ester and fatty acid, both at a concentration of 8 mM, we get a fatty acid production rate of about $0.65 \times 10^{-4}/s \times 8.0 \times 10^{-3}M = 5.2 \times 10^{-7} M/s$. Given this rate, τ_{double} is the time required to produce 8 mM more fatty acid from 8 mM picolinium ester concentration. Using $8 \times 10^{-3} M = \tau_{double} \times 5.2 \times 10^{-7} M/s$ yields $\tau_{double} = 1.54 \times 10^4 s$ or about 4.3 hours.

4 Simulation Exploration of Co-factor Replication

Non-enzymatic molecular replication was introduced in Section 2 based on lesion-induced DNA amplification (LIDA). This section explores LIDA for the protocellular information system through simulations of kinetic equations. The full kinetic system is introduced below in **Figure 10** and Eqs. (12)-(26) [37,38,57]. The parameters (rate constants) of the kinetic system are computed for a selection of DNA strands using established research on the thermodynamics of DNA structural motifs. Finally, LIDA-based simulations are performed for the 160 unique DNA strands (**Table 3**) in Appendix A already used for the CT simulations in Section 3 above.

4.1 The LIDA system for protocells

Alladin-Mustan et al. (2015) reported an exponential amplification of DNA replication under isothermal conditions at room temperature by introducing destabilizing lesions into the DNA duplex (LIDA or Lesion Induced DNA Amplification), such as abasic sites and mismatches [43]. These lesions decrease the free energy of the duplex's hydrogen bonds to the point where thermal fluctuations at room temperature can deliver the necessary energy to cause dehybridization. Furthermore, Alladin-Mustan et al. (2015) showed that the optimal temperature for DNA amplification can be tuned by the type and number of lesions added.

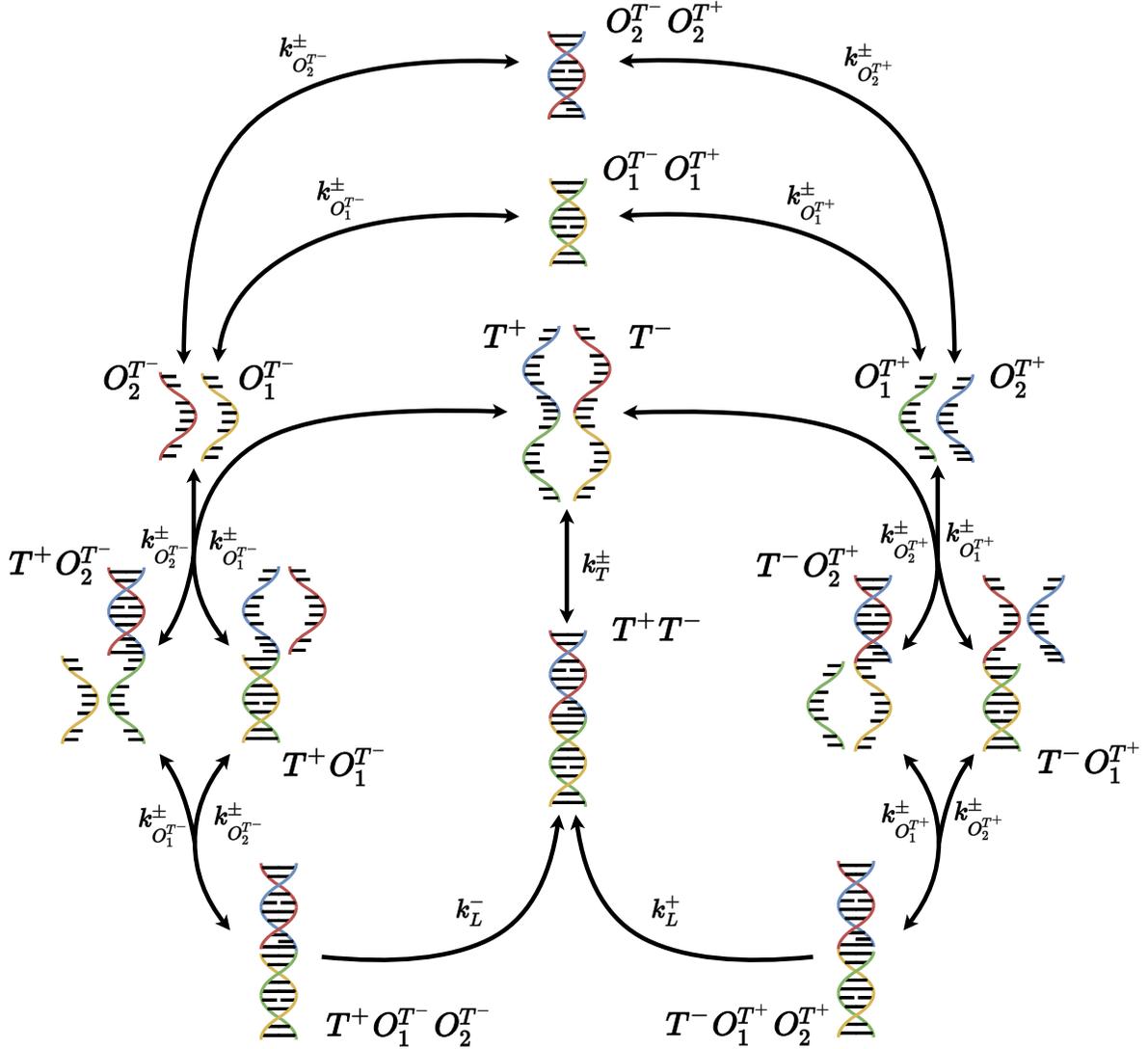
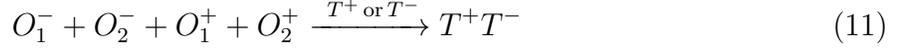


Figure 10: The LIDA system consists of four DNA single stranded oligomers (O_1^{T+} , O_2^{T+} , O_1^{T-} and O_2^{T-}) and two single stranded DNA templates (T^+ and T^-). Here, the reaction scheme of LIDA with O_1^{T+} , O_2^{T+} , O_1^{T-} and O_2^{T-} color coded as green, blue, yellow and red respectively, as well as the forward (k^+) and backward (k^-) reactions rates of all reactions. The oligomers O_1^{T+} and O_1^{T-} are near complementary (complementary except for mismatches and bulges; respectively illustrated by partially or fully unconnected nucleobases). Likewise, O_2^{T+} and O_2^{T-} are near complementary. These oligomers can thus form the short duplexes $O_2^{T-} O_2^{T+}$ and $O_1^{T-} O_1^{T+}$. Templates are longer single-stranded oligomers of the same sequence as two ligated short oligomers, i.e., ligating O_1^{T+} and O_2^{T+} forms T^+ and likewise ligating O_1^{T-} and O_2^{T-} forms T^- . The templates T^+ and T^- are thus also near complementary. The oligomers O_1^{\pm} and O_2^{\pm} can hybridize with the templates T^{\mp} forming four possible strands with sticky ends ($T^+ O_2^{T-}$, $T^+ O_1^{T-}$, $T^- O_1^{T+}$ and $T^- O_2^{T+}$). These four strands can further hybridize with oligomers that are near complementary to their sticky ends, forming the unligated strands $T^+ O_1^{T-} O_2^{T-}$ and $T^- O_1^{T+} O_2^{T+}$. A reaction can irreversibly ligate these strands into the DNA duplex $T^+ T^-$, which can dehybridize yielding new templates. The LIDA system thereby amplifies the concentration of templates while oligomers are present; in other words, LIDA makes DNA replication of a subset of strands possible without the complex mechanisms used in biological life.

The total reaction scheme of the LIDA system is shown in Figure 10, but can be summa-

rized by the reaction equation



meaning LIDA is the conversion of four short oligomers ($O_1^- + O_2^- + O_1^+ + O_2^+$) into the full DNA strand in duplex. This reaction is catalyzed by either simplex (T^+ or T^-) of the full DNA strand², and since the duplex and simplex forms can be converted into each other via hybridization/dehybridization, the total system is an autocatalytic production of double stranded DNA from four single-stranded oligomers.

It should be noted that the 5' end of the oligomers need to be imidazole activated for the ligation to occur, and if such hydrolysis occurs the deactivated oligomer is no longer able to take part in a ligation reaction. However, it is known from experiments [58] that the hydrolysis rate constant $k_{hyd-imp}$ of the imidazole activated oligomers is significantly lower than the ligation rate constant as well as the involved hybridization rates. Therefore, the imidazole hydrolysis process is not explicitly included in our LIDA simulations.

In previous experimental work [35] we have demonstrated that a picolil protection group can be attached to the 3' end of the oligomers, and that this protection group can be cleaved off using the same ruthenium-complex-based photo-activation process that generates fatty acids from picolin ester as discussed in previous Section 3. For simplicity, the details of this process are also not included in the LIDA simulation.

The kinetic equations of the LIDA system are shown in Equations (12)-(26). The equations are taken from [57] with new additional terms, due to the possibility that the near complementary oligomers form short duplexes ($O_1^{T+} O_1^{T-}$ and $O_2^{T+} O_2^{T-}$). This system of coupled differential equations is later used in Section 4.3 for *in silico* simulations of the LIDA system in order to obtain the replication dynamics of a selection of protocell co-factors.

$$\begin{aligned} \frac{d[O_1^{T+}]}{dt} = & k_{O_1^{T+}}^- [T^- O_1^{T+}] + k_{O_1^{T+}}^- [T^- O_1^{T+} O_2^{T+}] - k_{O_1^{T+}}^+ [T^-][O_1^{T+}] \\ & - k_{O_1^{T+}}^+ [T^- O_2^{T+}][O_1^{T+}] - k_{O_1^{T\pm}}^+ [O_1^{T-}][O_1^{T+}] + k_{O_1^{T\pm}}^- [O_1^{T+} O_1^{T-}] \end{aligned} \quad (12)$$

$$\begin{aligned} \frac{d[O_2^{T+}]}{dt} = & k_{O_2^{T+}}^- [T^- O_2^{T+}] + k_{O_2^{T+}}^- [T^- O_1^{T+} O_2^{T+}] - k_{O_2^{T+}}^+ [T^-][O_2^{T+}] \\ & - k_{O_2^{T+}}^+ [T^- O_1^{T+}][O_2^{T+}] - k_{O_2^{T\pm}}^+ [O_2^{T-}][O_2^{T+}] + k_{O_2^{T\pm}}^- [O_2^{T+} O_2^{T-}] \end{aligned} \quad (13)$$

²A non-catalyzed reaction via pseudo blunt-ended ligation may also occur, but much less frequent than the catalyzed reaction [43]. Furthermore, in the currently lab-achieved DNA amplification by LIDA, the system is also catalyzed by enzymes as noted in the Supporting Information for Alladin-Mustan et al. (2015) [43].

$$\begin{aligned} \frac{d[O_1^{T-}]}{dt} = & k_{O_1^-}^- [T^+ O_1^{T-}] + k_{O_1^-}^- [T^+ O_1^{T-} O_2^{T-}] - k_{O_1^-}^+ [T^+] [O_1^{T-}] \\ & - k_{O_1^-}^+ [T^+ O_2^{T-}] [O_1^{T-}] - k_{O_1^\pm}^+ [O_1^{T-}] [O_1^{T+}] + k_{O_1^\pm}^- [O_1^{T+} O_1^{T-}] \end{aligned} \quad (14)$$

$$\begin{aligned} \frac{d[O_2^{T-}]}{dt} = & k_{O_2^-}^- [T^+ O_2^{T-}] + k_{O_2^-}^- [T^+ O_1^{T-} O_2^{T-}] - k_{O_2^-}^+ [T^+] [O_2^{T-}] \\ & - k_{O_2^-}^+ [T^+ O_1^{T-}] [O_2^{T-}] - k_{O_2^\pm}^+ [O_2^{T-}] [O_2^{T+}] + k_{O_2^\pm}^- [O_2^{T+} O_2^{T-}] \end{aligned} \quad (15)$$

$$\begin{aligned} \frac{d[T^- O_1^{T+}]}{dt} = & k_{O_1^+}^+ [T^-] [O_1^{T+}] + k_{O_2^+}^- [T^- O_1^{T+} O_2^{T+}] - k_{O_1^+}^- [T^- O_1^{T+}] \\ & - k_{O_2^+}^+ [T^- O_1^{T+}] [O_2^{T+}] \end{aligned} \quad (16)$$

$$\begin{aligned} \frac{d[T^- O_2^{T+}]}{dt} = & k_{O_2^+}^+ [T^-] [O_2^{T+}] + k_{O_1^+}^- [T^- O_1^{T+} O_2^{T+}] - k_{O_2^+}^- [T^- O_2^{T+}] \\ & - k_{O_1^+}^+ [T^- O_2^{T+}] [O_1^{T+}] \end{aligned} \quad (17)$$

$$\begin{aligned} \frac{d[T^- O_1^{T+} O_2^{T+}]}{dt} = & k_{O_1^+}^+ [T^- O_2^{T+}] [O_1^{T+}] + k_{O_2^+}^+ [T^- O_1^{T+}] [O_2^{T+}] - k_{O_1^+}^- [T^- O_1^{T+} O_2^{T+}] \\ & - k_{O_2^+}^- [T^- O_1^{T+} O_2^{T+}] - k_L [T^- O_1^{T+} O_2^{T+}] \end{aligned} \quad (18)$$

$$\begin{aligned} \frac{d[T^+ T^-]}{dt} = & k_L [T^- O_1^{T+} O_2^{T+}] + k_L [T^+ O_1^{T-} O_2^{T-}] + k_T^+ [T^-] [T^+] \\ & - k_T^- [T^+ T^-] \end{aligned} \quad (19)$$

$$\begin{aligned} \frac{d[T^+ O_1^{T-}]}{dt} = & k_{O_1^-}^+ [T^+] [O_1^{T-}] + k_{O_2^-}^- [T^+ O_1^{T-} O_2^{T-}] - k_{O_1^-}^- [T^+ O_1^{T-}] \\ & - k_{O_2^-}^+ [T^+ O_1^{T-}] [O_2^{T-}] \end{aligned} \quad (20)$$

$$\begin{aligned} \frac{d[T^+ O_2^{T-}]}{dt} = & k_{O_2^-}^+ [T^+] [O_2^{T-}] + k_{O_1^-}^- [T^+ O_1^{T-} O_2^{T-}] - k_{O_2^-}^- [T^+ O_2^{T-}] \\ & - k_{O_1^-}^+ [T^+ O_2^{T-}] [O_1^{T-}] \end{aligned} \quad (21)$$

$$\begin{aligned} \frac{d[T^+ O_1^{T-} O_2^{T-}]}{dt} = & k_{O_1^-}^+ [T^+ O_2^{T-}] [O_1^{T-}] + k_{O_2^-}^+ [T^+ O_1^{T-}] [O_2^{T-}] - k_{O_1^-}^- [T^+ O_1^{T-} O_2^{T-}] \\ & - k_{O_2^-}^- [T^+ O_1^{T-} O_2^{T-}] - k_L [T^+ O_1^{T-} O_2^{T-}] \end{aligned} \quad (22)$$

$$\begin{aligned} \frac{d[T^-]}{dt} = & k_{O_1^+}^- [T^- O_1^{T+}] + k_{O_2^+}^- [T^- O_2^{T+}] + k_T^- [T^+ T^-] - k_{O_1^+}^+ [T^-] [O_1^{T+}] \\ & - k_{O_2^+}^+ [T^-] [O_2^{T+}] - k_T^+ [T^-] [T^+] \end{aligned} \quad (23)$$

$$\begin{aligned} \frac{d[T^+]}{dt} = & k_{O_1^-}^- [T^+ O_1^{T-}] + k_{O_2^-}^- [T^+ O_2^{T-}] + k_T^+ [T^+ T^-] - k_{O_1^-}^+ [T^+] [O_1^{T-}] \\ & - k_{O_2^-}^+ [T^+] [O_2^{T-}] - k_T^- [T^+] [T^-] \end{aligned} \quad (24)$$

$$\frac{d[O_1^{T+} O_1^{T-}]}{dt} = k_{O_1^\pm}^+ [O_1^{T-}] [O_1^{T+}] - k_{O_1^\pm}^- [O_1^{T+} O_1^{T-}] \quad (25)$$

$$\frac{d[O_2^{T^+} O_2^{T^-}]}{dt} = k_{O_2^{T^\pm}}^+ [O_2^{T^-}] [O_2^{T^+}] - k_{O_2^{T^\pm}}^- [O_2^{T^+} O_2^{T^-}] \quad (26)$$

4.2 Kinetic rates in LIDA

All reactions in the LIDA kinetic system are of the two types:



except for the irreversible ligation reaction, where the observed rate from [35] is used. Equilibrium thermodynamics gives a relation between the standard Gibbs free energy ΔG° and the equilibrium constant K :

$$\Delta G^\circ = -RT \ln K, \quad (28)$$

where $R = 1.987 \text{ cal} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ is the gas constant and T is the temperature in Kelvin. Using the definition of the equilibrium constant $K = k^+/k^-$, we can obtain:

$$K = \frac{k^+}{k^-} = \exp\left(-\frac{\Delta G^\circ}{RT}\right) \Rightarrow k^- = k^+ \exp\left(\frac{\Delta G^\circ}{RT}\right) \quad (29)$$

Thus, if we know the on-rate (forward rate constant) k^+ , the temperature T , and the standard Gibbs free energy ΔG° , we can compute the off-rate (reverse rate constant) k^- . For simplicity, we assume all on-rates to be $2 \cdot 10^7 \text{ s}^{-1}$ [59], while ΔG° is computed using the method and parameters introduced by SantaLucia and Hicks (2004) [54]. From this, we compute off-rates ($k_{O_1^{T^-}}^-$, $k_{O_1^{T^+}}^-$, $k_{O_2^{T^-}}^-$, $k_{O_2^{T^+}}^-$ and k_T^-) for LIDA simulations.

As mentioned, protocellular co-factors contain destabilizing mismatches to obtain faster off-rates, and specifically, the 160 investigated strands (see Table 3) contain G/A and G/T mismatches. Mismatch contributions to ΔG° at 37°C are listed in Table 2 in [54]. For our LIDA simulations (in which the temperature is set to 26°C), we must instead use ΔH° and ΔS° to compute $\Delta G_{26^\circ}^\circ$ for G/T and G/A mismatches, which can be found in [60,61].

The investigated protocell genomes also contain a highly destabilizing bulge, which forms near the center of the templates in duplex (T^+T^-), replacing the abasic site as the main destabilizing element in the original LIDA system [43]. This is required to obtain faster dehybridization rates for the T^+T^- duplex (k_T^-). SantaLucia and Hicks (2004) state that bulge loops of size 1 contribute $\Delta G_{\text{bulge},37^\circ}^\circ = 4 \text{ kcal} \cdot \text{mol}^{-1}$. A negative correction should be added to obtain $\Delta G_{\text{bulge},26^\circ}^\circ$, but since we do not know the size of this correction, we simply use $\Delta G_{\text{bulge},37^\circ}^\circ = 4 \text{ kcal} \cdot \text{mol}^{-1}$ in our simulations.

Dangling ends (DEs) also contribute to the hybridization free energy. Including DE contributions can have a significant effect on the LIDA replication dynamics, particularly when computing off-rates for single oligomer/templates duplexes ($T^-O_1^{T^+}$, $T^-O_2^{T^+}$, $T^+O_1^{T^-}$ and $T^+O_2^{T^-}$). DEs are predominantly stabilizing, while contributions from single nucleotide range from +0.48 to $-0.96 \text{ kcal} \cdot \text{mol}^{-1}$ depending on sequence and orientation (5'-3' or 3'-5') [54]. Nearly all of the DE contribution comes from the first DE, with additional nucleotides contributing less than $0.2 \text{ kcal} \cdot \text{mol}^{-1}$ [54].

For our implementation, we included a single correcting factor to the off-rates in the terms of dehybridization of the short oligomers from the strands with dangling ends, i.e., $T^-O_1^{T^+}$, $T^-O_2^{T^+}$, $T^+O_1^{T^-}$ and $T^+O_2^{T^-}$. SantaLucia and Hicks (2004) state that the average 5'-dangling end contributes $\Delta G_{5'}^\circ = -0.45 \text{ kcal} \cdot \text{mol}^{-1}$, while the average 3'-dangling end contributes $\Delta G_{3'}^\circ = -0.29 \text{ kcal} \cdot \text{mol}^{-1}$. We therefore chose the average of $\Delta G_{5'}^\circ$ and $\Delta G_{3'}^\circ$ as our correcting factor to the hybridization free energy: $\Delta G_{DE}^\circ = (-0.45 - 0.29)/2 = -0.37 \text{ kcal} \cdot \text{mol}^{-1}$. Though a more rigorous calculation of the sequence dependence of DE contribution is possible, we use this single correcting factor to test the sensitivity of the LIDA kinetic system to DE contributions. When comparing LIDA simulations (as specified in Section 4.3) including and excluding DE contributions, we observe a decrease in replication times of protocellular co-factors via LIDA when DE contributions are included. A typical LIDA simulation is shown in Fig. 11.

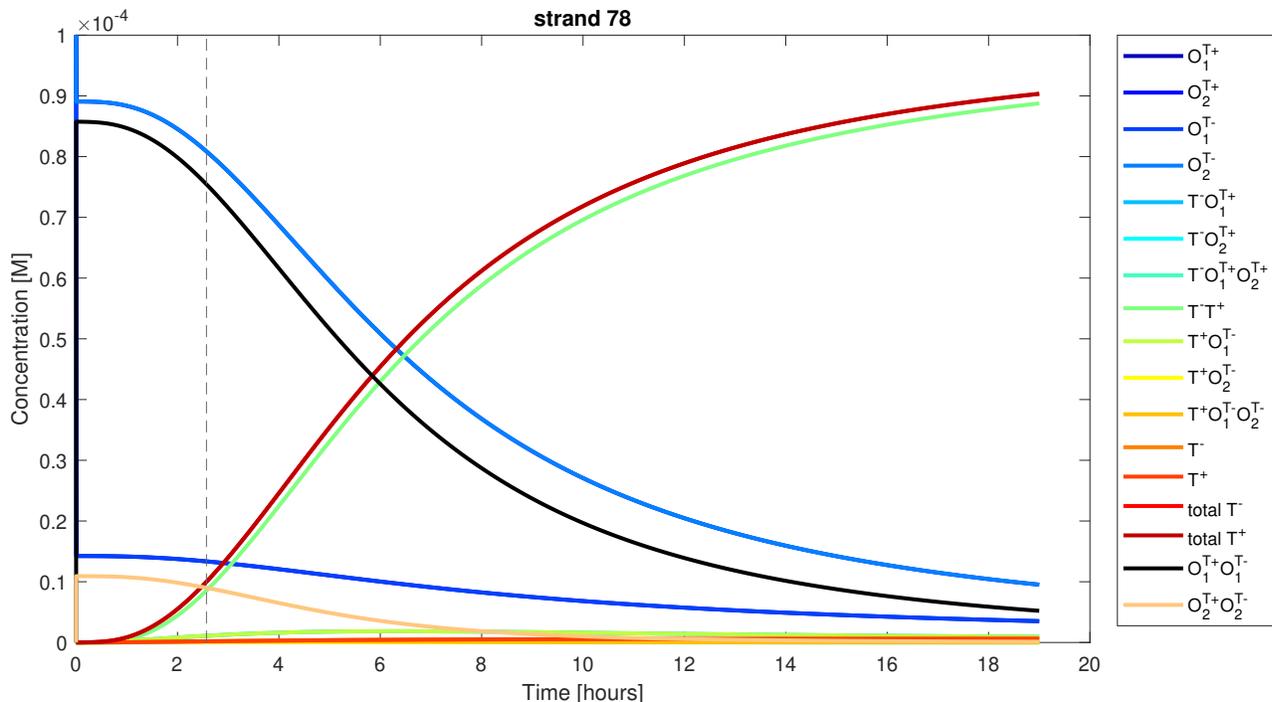


Figure 11: Typical LIDA simulation where the graph color corresponds to the molecular complex indicated at the right side of the figure. Note how the two oligomer duplexes are formed at the onset, with $O_1^{T+}O_1^{T-}$ (black line) in significantly higher concentration (higher hybridization energies) than the $O_2^{T+}O_2^{T-}$ (beige line) duplex. These duplex concentrations thereafter decrease as the concentration of the total number of full plus-and-minus strands increases exponentially (red line) where they are mostly in a duplex configuration (green line, right below the red line). The light blue curve (above the black line) shows the kinetics of the free O_2^{T+} and O_2^{T-} oligomers, while the dark blue curve (above the beige line) shows the kinetics of the free O_1^{T+} and O_1^{T-} oligomers. The many intermediary template-oligomer complexes as well as the free plus and minus templates are at low concentrations throughout the simulation and therefore difficult to distinguish at this concentration resolution.

We note that there are inconsistencies between the thermodynamic parameters computed using the method by SantaLucia and Hicks and the thermodynamic parameters computed by NUPACK [57]. All thermodynamic parameters used to compute off-rates for LIDA simulations in this paper are listed in Table 2.

NN base pair	ΔH° [kcal·mol ⁻¹]	ΔS° [cal·mol ⁻¹ ·K ⁻¹]
AA/TT	-7.6	-21.3
TT/AA	-7.6	-21.3
AT/TA	-7.2	-20.4
TA/AT	-7.2	-21.3
GT/CA	-8.4	-22.4
GA/CT	-8.2	-22.4
AG/TC	-8.2	-22.4
TG/AC	-8.4	-22.4

Mismatch type	ΔH° [kcal·mol ⁻¹]	ΔS° [cal·mol ⁻¹ ·K ⁻¹]
AG/TA	-0.7	-2.3
AG/TT	1	0.9
AT/TG	-2.5	-8.3
AA/TG	-0.6	-2.3

Lesion type	$\Delta G_{37^\circ}^\circ$ [kcal·mol ⁻¹]
bulge loop of size 1	4

Table 2: The thermodynamic parameters used to compute free energies of DNA hybridization for LIDA simulations. All parameters have been found in [54, 60, 61] and is denoted using the same notation, where e.g. AA/TT corresponds to 5'-AA-3' hybridized with 3'-TT-5'.

4.3 LIDA simulations for a selection of strands

By using DNA strand-specific kinetic off-rates from thermodynamic parameters and the kinetic equations of LIDA, we simulate co-factor replication for the same 160 DNA strands that we consider in Section 3.2 when studying charge transport. This is done by splitting all strands into oligomers of equal length (9 bp or 3 substrands). For each strand, the hybridization free energies, Eq. (28), are computed for the short duplexes ($O_1^{T-}O_1^{T+}$ and $O_2^{T-}O_2^{T+}$) and the templates in duplex (T^+T^-) using the method of SantaLucia and Hicks [54]. Note that the bulge only contributes to the free energy for T^+T^- . The thermodynamic contribution of 8-oxo-G is unknown, but Gasper and Schuster (1997) note that replacement of G by 8-oxo-G has little or no effect on the global duplex structure [51]. We therefore assume that 8-oxo-G can be treated as a G in calculations of hybridization free energy.

Inserting these free energies into Eq. (29) with $T = 26^\circ\text{C}$ yields off-rates, which can be implemented in the LIDA kinetic system, Eqs. (12)-(26). We assume that $k_{O_1^{T-}}^- = k_{O_1^{T+}}^-$ and $k_{O_2^{T-}}^- = k_{O_2^{T+}}^-$, although there may be discrepancies. The resulting kinetic systems are simulated using the numerical solver `ode15s` in MATLAB R2021b. If the simulation time reaches 100 hours, the simulation is stopped.

We estimate the replication rate constant k_{rep} by (i) measuring the time $\tau_{10\%}$ at which DNA

amplification has reached 10% of completion ($= 10 \times 10^{-6}$ M) and (ii) assuming that the initial growth phase can be approximated as exponential. Then k_{rep} can be estimated from 10×10^{-6} M $= 14 \times 10^{-9}$ M $\times \exp(k_{rep} \times \tau_{10\%})$ as $k_{rep} = \ln[(10/14) \times 10^3]/\tau_{10\%}$. The doubling or replication time τ_{rep} can be estimated from $A \exp(k_{rep} \times \tau_{rep}) = 2A$ so $\tau_{rep} = \ln 2/k_{rep}$. **Figure 12** shows a histogram of the replication rates obtained for the 160 strands.

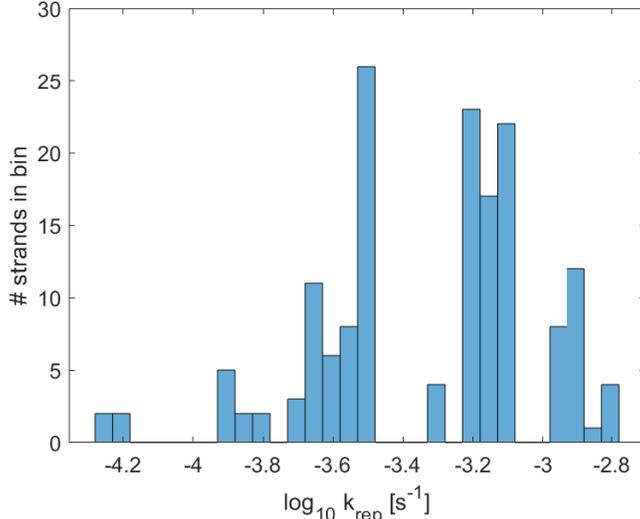


Figure 12: Histogram showing the distribution of LIDA simulated replication rates k_{rep} for the set of 160 strands. Initial concentrations used in the LIDA simulations for the 160 strands are $[O_1^{T+}] = [O_1^{T-}] = [O_2^{T+}] = [O_2^{T-}] = 100 \mu\text{M}$ and $[T^-] = 14 \text{ nM}$ with all other initial concentrations equal to zero M. We find typical doubling times for k_{rep} between $10^{-4}/\text{s}$ and $10^{-3}/\text{s}$ where the corresponding doubling times are $\tau_{rep} = 6.70 \times 10^3 \text{ s}$ and $6.70 \times 10^2 \text{ s}$ respectively.

The estimated replication rate constant k_{rep} data in **Figure 12** show multiple semi-distinct groupings, with typical doubling times ranging between about 10 minutes and a couple of hours, with the slowest about 3.8 hours. Note that this is faster than the estimated vesicle doubling time of 4.3, we found in Section 3.

Finally, we need to check the above estimated replication rate constants with the imidazole hydrolysis rate constant $k_{hyd-imp}$ that is measured to be between $4.78 \times 10^{-6}/\text{s}$ and $1.04 \times 10^{-5}/\text{s}$ at room temperature and depending on the ion concentration [58]. DNA sequences with an estimated overall replication rate constant below $k_{hyd-imp}$ are not viable as the replication process will be inhibited by the decay of activated oligomers.

Since most of our estimated replication rate constants satisfy $k_{rep} > 10^{-4}/\text{s}$, which is an order of magnitude larger than the largest reported hydrolysis rate constants $k_{hyd-imp} \sim 10^{-5}/\text{s}$, as a first approximation we can assume that a protocellular replication is viable if we require $k_{rep} > 10^{-4}/\text{s}$.

5 Connecting results from charge transport and replication simulations

The data on the replication rate and CT rates from the simulations performed in Sections 3.2 and 4.3 provide us with information about how the protocell fitness depends on the co-factor: Assuming everything else being equal, higher replication and/or CT rates mean higher fitness.

Obviously, a lower bound must exist for both the co-factor charge transfer and replication rates if the protocellular system should be able to grow and survive. If either the CT rate or the replication rate are too slow, a variety of degradation processes would exceed the production processes and the protocellular system would either not be able to form or it would disintegrate. Recall that from our analysis in Section 3.2, we require that $k_{CT} > 50/\text{s}$ to ensure that the CT rate constant is well above the rate-limiting photo-activation rate constant. From our analysis in Section 4.3, we require that $k_{rep} > 10^{-4}/\text{s}$ to ensure that the replication rate constant is well above the associated hydrolysis rates. These conditions ensure that charge-transport and replication processes can support viable protocells. Also, recall the discussion in Section 2, where for simplicity we required that 8-oxo-G is located above the ligation site, which disqualifies strand numbers 65-80 and 113 to 160 (see Table 3 for details). Thus, 95 out of the 160 strands satisfy this condition. In Figure 13 we show the replication rate k_{rep} versus the CT rate k_{CT} for all 160 examined strands, where the ‘viable area’ is indicated. In Table 3, the 160 strands are listed where their replication and charge transfer properties are encoded using similar colors as used in Figure 13.

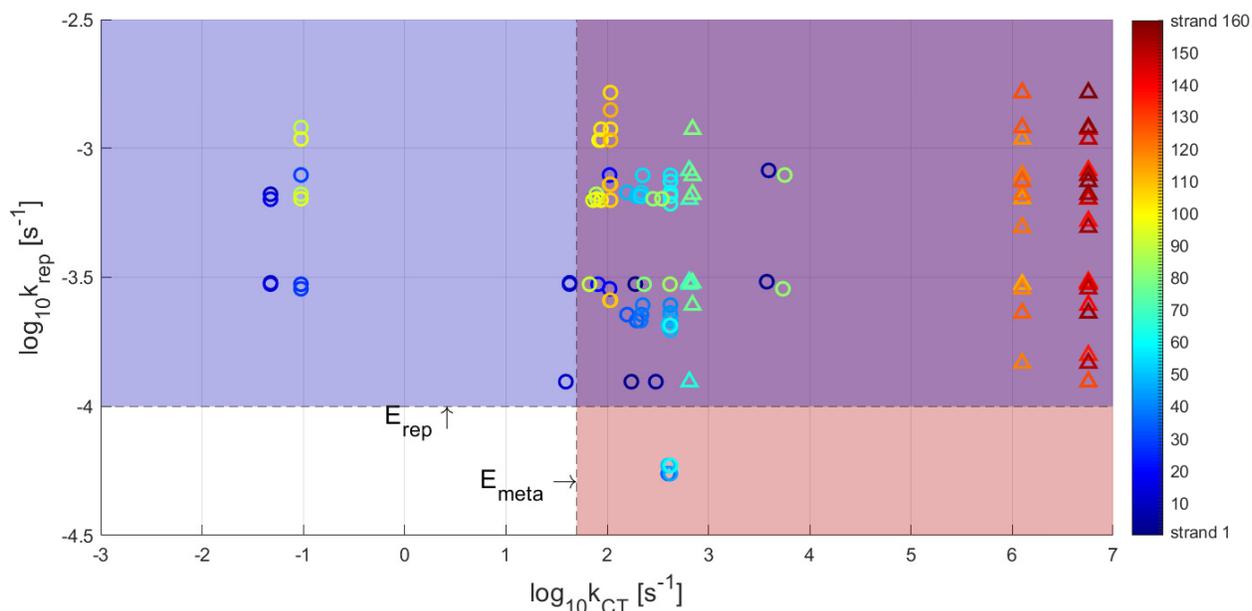


Figure 13: The replication rate k_{rep} versus the CT rate k_{CT} for all 160 examined strands (see Table 3 for a listing of all strands). Color bar indicates which string corresponds to which data point; rings indicate that oxoG is above the ligation site, while triangles indicate that oxoG is below the ligation site. Everything else being equal, both the replication rate and the charge transfer rate have to be above some minimal value to ensure protocellular survival. This is indicated by the E_{meta} and E_{rep} values and their respectively pale red and pale blue areas. Thus, the purple area indicates both viable metabolic and replication rates. See text for details.

It should be emphasized that as we combine the results from the CT and LIDA simulations, no interaction dynamics are taken into consideration. This includes interactions between the replication and the charge transfer processes, as well as the feeding, growth, and container division processes. Further, our simulations assume a constant environment (temperature, pH, salt, etc.). Thus, a combination of the results found under the CT and LIDA simulation can only be viewed as a simplified representation of what a full systems simulation would have given. This means that the viable area shown in the reaction rate diagram in Figure 13 should be viewed as a simplified approximation.

Strand	Substrand sequence						
1	1 1 6 1 1 5	41	1 3 6 2 5 1	81	2 1 6 1 1 5	121	2 5 6 1 4 1
2	1 1 6 1 2 5	42	1 3 6 2 5 2	82	2 1 6 1 2 5	122	2 5 6 1 4 2
3	1 1 6 1 3 5	43	1 3 6 5 1 1	83	2 1 6 1 3 5	123	2 5 6 2 1 1
4	1 1 6 1 4 5	44	1 3 6 5 1 2	84	2 1 6 1 4 5	124	2 5 6 2 1 2
5	1 1 6 1 5 1	45	1 3 6 5 1 3	85	2 1 6 1 5 1	125	2 5 6 2 1 3
6	1 1 6 1 5 2	46	1 3 6 5 1 4	86	2 1 6 1 5 2	126	2 5 6 2 1 4
7	1 1 6 2 1 5	47	1 3 6 5 2 1	87	2 1 6 2 1 5	127	2 5 6 2 2 1
8	1 1 6 2 2 5	48	1 3 6 5 2 2	88	2 1 6 2 2 5	128	2 5 6 2 2 2
9	1 1 6 2 5 1	49	1 4 6 1 1 5	89	2 1 6 2 5 1	129	5 1 6 1 1 1
10	1 1 6 2 5 2	50	1 4 6 1 2 5	90	2 1 6 2 5 2	130	5 1 6 1 1 2
11	1 1 6 5 1 1	51	1 4 6 1 3 5	91	2 1 6 5 1 1	131	5 1 6 1 1 3
12	1 1 6 5 1 2	52	1 4 6 1 4 5	92	2 1 6 5 1 2	132	5 1 6 1 1 4
13	1 1 6 5 1 3	53	1 4 6 1 5 1	93	2 1 6 5 1 3	133	5 1 6 1 2 1
14	1 1 6 5 1 4	54	1 4 6 1 5 2	94	2 1 6 5 1 4	134	5 1 6 1 2 2
15	1 1 6 5 2 1	55	1 4 6 2 1 5	95	2 1 6 5 2 1	135	5 1 6 1 3 1
16	1 1 6 5 2 2	56	1 4 6 2 2 5	96	2 1 6 5 2 2	136	5 1 6 1 3 2
17	1 2 6 1 1 5	57	1 4 6 2 5 1	97	2 2 6 1 1 5	137	5 1 6 1 4 1
18	1 2 6 1 2 5	58	1 4 6 2 5 2	98	2 2 6 1 2 5	138	5 1 6 1 4 2
19	1 2 6 1 3 5	59	1 4 6 5 1 1	99	2 2 6 1 3 5	139	5 1 6 2 1 1
20	1 2 6 1 4 5	60	1 4 6 5 1 2	100	2 2 6 1 4 5	140	5 1 6 2 1 2
21	1 2 6 1 5 1	61	1 4 6 5 1 3	101	2 2 6 1 5 1	141	5 1 6 2 1 3
22	1 2 6 1 5 2	62	1 4 6 5 1 4	102	2 2 6 1 5 2	142	5 1 6 2 1 4
23	1 2 6 2 1 5	63	1 4 6 5 2 1	103	2 2 6 2 1 5	143	5 1 6 2 2 1
24	1 2 6 2 2 5	64	1 4 6 5 2 2	104	2 2 6 2 2 5	144	5 1 6 2 2 2
25	1 2 6 2 5 1	65	1 5 6 1 1 1	105	2 2 6 2 5 1	145	5 2 6 1 1 1
26	1 2 6 2 5 2	66	1 5 6 1 1 2	106	2 2 6 2 5 2	146	5 2 6 1 1 2
27	1 2 6 5 1 1	67	1 5 6 1 1 3	107	2 2 6 5 1 1	147	5 2 6 1 1 3
28	1 2 6 5 1 2	68	1 5 6 1 1 4	108	2 2 6 5 1 2	148	5 2 6 1 1 4
29	1 2 6 5 1 3	69	1 5 6 1 2 1	109	2 2 6 5 1 3	149	5 2 6 1 2 1
30	1 2 6 5 1 4	70	1 5 6 1 2 2	110	2 2 6 5 1 4	150	5 2 6 1 2 2
31	1 2 6 5 2 1	71	1 5 6 1 3 1	111	2 2 6 5 2 1	151	5 2 6 1 3 1
32	1 2 6 5 2 2	72	1 5 6 1 3 2	112	2 2 6 5 2 2	152	5 2 6 1 3 2
33	1 3 6 1 1 5	73	1 5 6 1 4 1	113	2 5 6 1 1 1	153	5 2 6 1 4 1
34	1 3 6 1 2 5	74	1 5 6 1 4 2	114	2 5 6 1 1 2	154	5 2 6 1 4 2
35	1 3 6 1 3 5	75	1 5 6 2 1 1	115	2 5 6 1 1 3	155	5 2 6 2 1 1
36	1 3 6 1 4 5	76	1 5 6 2 1 2	116	2 5 6 1 1 4	156	5 2 6 2 1 2
37	1 3 6 1 5 1	77	1 5 6 2 1 3	117	2 5 6 1 2 1	157	5 2 6 2 1 3
38	1 3 6 1 5 2	78	1 5 6 2 1 4	118	2 5 6 1 2 2	158	5 2 6 2 1 4
39	1 3 6 2 1 5	79	1 5 6 2 2 1	119	2 5 6 1 3 1	159	5 2 6 2 2 1
40	1 3 6 2 2 5	80	1 5 6 2 2 2	120	2 5 6 1 3 2	160	5 2 6 2 2 2

Table 3: All 160 strands used for simulations and their substrand sequences. The nucleotide sequences of different substrands are shown in Equation 3-8. Pale red cells contain strands that satisfy the condition for sufficiently fast CT and pale blue cells contain strands that satisfies the condition for sufficiently fast replication (see Section 5.1.2). Purple cells contain strands that satisfy both of the aforementioned conditions, just as in Figure 13. Black text signifies a strand sequence containing an 8-oxo-G after the ligation site (rings in Figure 13) while grey text means the 8-oxo-G is before the central bulge (triangles in Figure 13). Red cells contain strands for which either the LIDA or CT simulation did not finish within the allotted time. Note that the number of viable strands, indicated by purple cells and black text, is 63.

5.1 Functional characterization of co-factors

5.1.1 Direct rate combination

To further explore the viability transition discussed above, we can in principle rank the combinatorial co-factors $F(k_{rep}, k_{CT}, \dots)$ such that their combined replication and charge transfer (plus subsequent fatty acid production) rates are increasing along an axis. Again note that a reliable combined co-factor 'fitness' ranking cannot yet be estimated without a complete protocell-environment simulation. However, we can still study co-factor rankings based on the results of the CT and LIDA simulations as they are of theoretical interest.

A simplistic method to approximate the combined fitness impact of the charge transfer (CT) and the replication rates is to simply multiply the two rates. This can be done by combining the replication rate data and the CT rate data as the sum of the \log_{10} of replication rate and the CT rate for each strand. There are, however, problems with a direct multiplication. Rocheleau et al. (2007) [29] studied a similar protocellular system that combined information replication and container growth. They showed that due to non-trivial couplings between interacting processes, the combined protocellular growth factor is proportional to $[k_{rep}^2 k_{CT}]^{1/3}$ rather than $(k_{rep} k_{CT})$. This is shown when the coupled equations (4.1)-(4.6) in [29] are solved analytically. Although the coupling between our charge transfer rate and our co-factor replication rate is different from the coupled aggregated template and container replication rates discussed in [29], the two systems are closely related and we believe that the results from [29] provide a reasonable ansatz for a combined fitness estimate given our current knowledge of these systems. This fitness estimate is shown in [Figure 14](#), which can be interpreted as a simple approximation for evaluating the fitness of a protocellular co-factor. Note, however, that the growth law found in [29] is derived under restricted assumptions and defining a more realistic fitness coefficient would require an expanded and more detailed model that is outside the scope of this work.

The sorted fitness coefficient data shows three groups similar to the clustered CT rate distribution, with added smoothness arising from the more continuous replication rate distribution. The three groups are mainly explained by the same factors as the groups from the CT rate data, recall [Section 3](#) and [Figure 8](#), as the CT rates vary across more orders of magnitude than the replication rates.

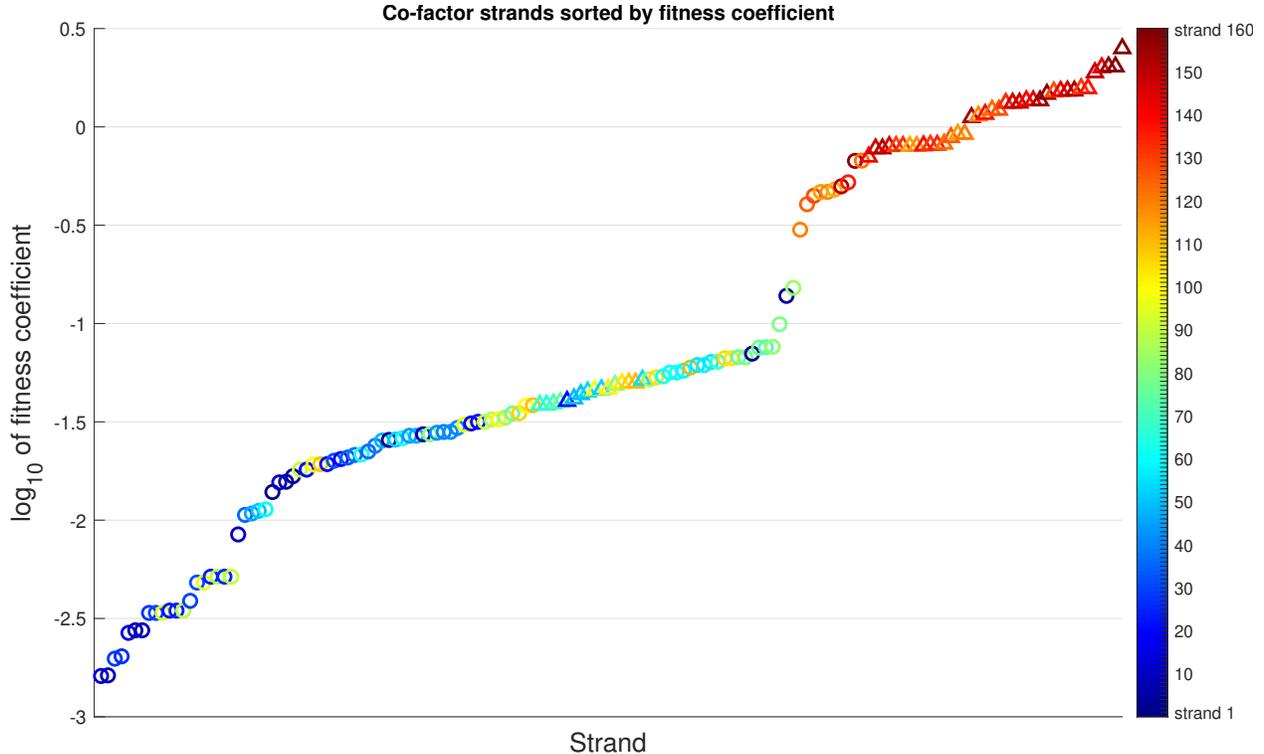


Figure 14: Ansatz for a combined metabolic and replication growth rate from [29] depicted as the sorted sum of $\frac{2}{3} \log_{10}[k_{rep}] + \frac{1}{3} \log_{10}[k_{CT}]$. Note the three growth rate regimes (blue to blue/green to red) dominated by the clustered charge transport rates. Also note that no metabolism-replication interaction are included in our simulations, so the graph is a simple (superposition) approximation of the protocellular fitness based on co-factor for the 160 examined strands. The color bar shows which strand corresponds to which data point; rings indicate that oxoG is above the ligation site, while and triangles indicate the oxoG is located below the ligation site. See text for details.

5.1.2 Functional information

An alternative way to explore the viability of the presented protocellular system is to estimate the functional information of the involved co-factors.

The concept of functional information was first introduced by Szostak (2003) to define and quantify the information content of biopolymer sequences [40]. Hazen et al. (2007) later used functional information as a more general measure of complexity [41]. Hazen et al. (2007) argue that common for all complex systems is the ability to alter their environment in one or more ways, which they refer to as functions. For a system with a specified function (x) and a set of configurations M , each configuration $m \in M$ of the system has a “degree of function” $E_x(m)$ defined by its ability to perform the function x . For a specific value of E_x , only a subset of the total configurations has the same-or-higher degree of function, $E_x(m) \geq E_x$. In mathematical terms, the size of this set is given by $\omega = |\{m \in M : E_x(m) \geq E_x\}|$. With the size of the total set of configurations $\Omega = |M|$, we can define the probability of a random

configuration having degree of function $E_x(m) \geq E_x$ as

$$P_x(E_x) = \frac{\omega}{\Omega}. \quad (30)$$

Functional information is then defined as

$$I[E_x] = -\log_2 P_x(E_x) = -\log_2 \left(\frac{\omega}{\Omega} \right). \quad (31)$$

As Hazen et al. (2007) concludes, a rigorous analysis of the functional information of a system with respect to a specified function x requires knowledge of two attributes: (i) all possible configurations M of the system and (ii) the degree of function x for every configuration [41]. The difficulty of obtaining knowledge of these two attributes clearly scales with the complexity of the system.

The concept of a function obviously has a connection to the concept of evolutionary fitness, as functions contribute to the fitness of an evolving entity. This is true for systems at multiple levels of complexity; e.g., the degree of function of a metabolic pathway of a resource molecule utilized by a microorganism, or the degree of function of the beak shape of a finch allowing consumption of a specific food source. In protocellular populations, fitness can be decomposed into two main functions: (i) metabolic function, e.g., the metabolic rate of lipid precursors into usable lipids (E_{meta}), which is regulated by the rate of charge transport through co-factors and (ii) replicative function, e.g., the replication rate of co-factors through LIDA (E_{rep}). The distribution of degrees of these functions in protocell populations therefore gives rise to a fitness distribution, and thus selection dynamics. As mentioned, a full analysis of these functions of the protocell information system would require complete knowledge of the possible configurations $m \in M_{proto}$ and their corresponding degrees of function $E_{meta}(m)$ and $E_{rep}(m)$.

M_{proto} is hard to define. An upper bound could be every possible configuration of the combinatorial co-factor, i.e., every possible DNA duplex of length 18 bp which can be formed using A, T, G, C and 8-oxo-G while also allowing mismatches, bulges and variations of oligomer lengths. The size Ω_{proto} of the set M_{proto} is then clearly a very large number; too big to ever permit a brute-force analysis of the functional information.

Our investigation of protocellular functional information is limited to the set of sequences used for which CT and LIDA simulations have been performed. As discussed above, based on the results from Sections 3 and 4, we calculated that minimum values of $k_{meta} > 50/s$ and $k_{rep} > 10^{-4}/s$ are needed to support a viable protocells, so we can now estimate the probabilities of randomly choosing a strand from the 160 tested strands that satisfies each of these conditions.

$$P(k_{CT} > 50/s) = \frac{[\# \text{ strands that satisfy } k_{CT} > 50/s]}{[\text{total } \# \text{ strands}]} \quad (32)$$

$$P(k_{rep} > 10^{-4}/s) = \frac{[\# \text{ strands that satisfy } k_{rep} > 10^{-4}/s]}{[\text{total } \# \text{ strands}]} \quad (33)$$

$$P(k_{CT} > 50/s \ \& \ k_{rep} > 10^{-4}/s) = \frac{[\# \text{ strands that satisfy } k_{CT} > 50/s \ \& \ k_{rep} > 10^{-4}/s]}{[\text{total } \# \text{ strands}]} \quad (34)$$

Needed compositional functional information in strands for viable protocells:

$$I(\text{CT is good}) = -\ln[P(k_{CT} > 50/s)] \quad (35)$$

$$I(\text{rep is good}) = -\ln[P(k_{rep} > 10^{-4}/s)] \quad (36)$$

$$I(\text{CT \& rep is good}) = -\ln[P(k_{CT} > 50/s \ \& \ k_{rep} > 10^{-4}/s)]. \quad (37)$$

Table 3 indicates the total number of protocells with co-factor functions greater than $E_{meta}(m) \geq 50/s$ and $E_{rep}(m) \geq 10^{-4}/s$ yields $\omega_{meta+rep} = 63$. We can now compute the total functional information needed for a viable protocell if a co-factor strand is randomly picked from from the ensemble of 160 strands,

$$I[E_{meta+rep}] = -\log_2(F(E_{meta+rep})) \quad (38)$$

$$= -\log_2\left(\frac{\omega_{meta+rep}}{\Omega_{proto}}\right) = -\log_2\left(\frac{63}{160}\right) \approx 1.34 \text{ bits.} \quad (39)$$

Thus, 1.34 bits is the estimated functional information required for a protocell to be viable when its co-factor is randomly drawn from the limited available set of 160 co-factors.

We can also calculate the corresponding functional information of the co-factors as they are ranked according to their metabolic and replication rate constants, which is shown in **Figure 15**.

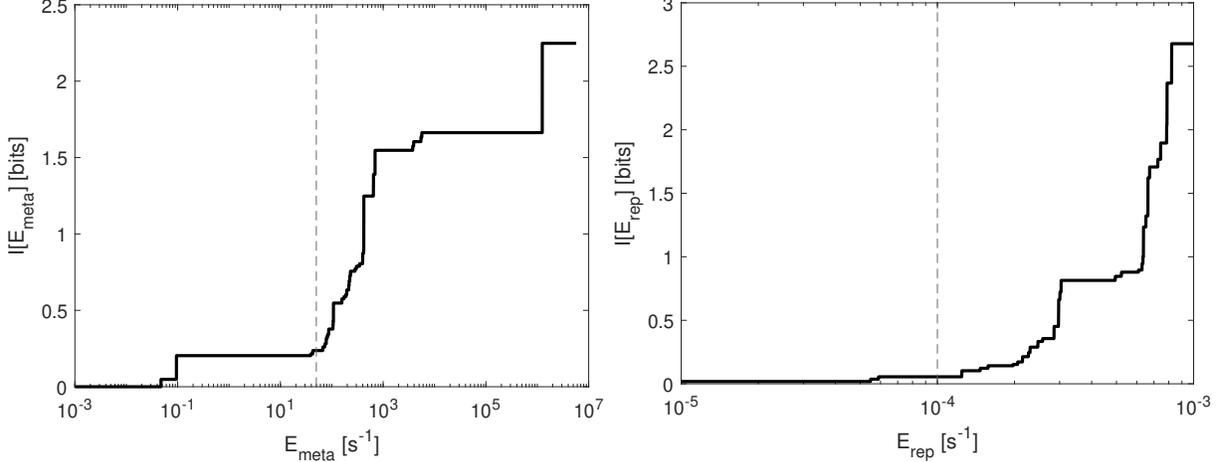


Figure 15: (Left) The functional information $I[E_{meta}]$ of co-factors as a function of E_{meta} and (Right) $I[E_{rep}]$ as a function of E_{rep} . Both show an increase in functional information with E_x as fewer co-factors have degrees of function $\geq E_x$. The vertical dashed lines indicate $E_{meta} = 50/s$ and $E_{rep} = 10^{-4}/s$ respectively. Note that the maximum functional information differs for the two functions, as the number of identical fastest replicators and charge-transfer strands differs. For instance, for the charge transfer function 32 strands have an identical charge transfer rate of about $4 \times 10^6/s$, yielding about 2.32 bits.

In **Figure 15**, functional information of both charge transport and replication reflect, in different ways, the same qualitative shape of the estimated protocellular growth rates shown in **Figure 14**. From **Table 3**, we see that the CT threshold $k_{CT} > 50/s$ (and 8-oxo-G above the ligation site) is satisfied for 67 strands, thus $\omega_{meta} \simeq 67$, while the replication threshold $k_{rep} > 10^{-4}/s$ (and 8-oxo-G above ligation site) is satisfied for 86 strands, thus $\omega_{rep} \simeq 86$. This corresponds to functional information thresholds of 1.26 bits and 0.89 bits for charge transport and replication functions, respectively. Note as a reference that the theoretical maximal possible functional information for this co-factor system is $-\log_2(1/160)$ bits $\simeq 7.30$ bits.

6 Discussion

The presented work consists of four main parts: (i) conceptual discussion of the protocellular functionalities and design; (ii) simulations of DNA charge transport between Ru-C and oxoG; (iii) simulations of lesion-induced DNA amplification; and (iv) a combination of the results from (ii) and (iii). In this section, we critically discuss the applied assumptions and approximations.

About (i): In Section 2 we present and discuss a conceptual picture of our protocellular design and functionalities, as a background for our later analysis of the two critical co-factor properties: replication and charge transfer. Our protocell system is also discussed in [27, 29, 33, 35] [36], while alternative protocellular designs are presented and discussed

in [1, 2, 8, 9, 12, 30].

About (ii): See Section 3. It is known that certain Ru-C merges into the DNA base stack (intercalation) and is able to oxidize guanines after photoexcitation [48, 49]. To induce this intercalation the Ru-C and DNA duplex are usually tethered. Without direct tethering a similar intercalation should be achieved by attaching both the Ru-C and two of the DNA duplex to hydrophilic anchors via carbon tethers. The goal is to only have charge transfer between the RuC and the oxoG through a duplex, such that the DNA replication process that creates more duplexes positively influence the fatty acid production process, thus establishing a catalytic coupling between the two.

Our choice of kinetic modeling for the CT simulations is based on its low computational cost. More refined methods, e.g., full quantum-mechanical simulations, would have drastically increased the computational cost of the simulations and thereby also drastically limited the number of DNA co-factors that could be simulated in the allotted time. Thus, it would not have been possible to compare the CT capabilities of multiple co-factor sequences.

To use ODEs means that basepair-to-basepair CT rates have to be known for each base pair CT interaction. Thus, we consulted the literature where such rates are reported, either from direct measurements or from *ab initio* calculations. As it turns out not all base pair CT combinations rates are listed in the available literature, we could only simulate a limited set of duplex strings. Also, several CT rates had to be estimated by the authors as no previous research was found (e.g., on CT rates through bulges).

It is not known to the authors how feasible super exchange (SE) transitions are through disruptions of the DNA π stack such as mismatches, bulges, and possibly 8-oxo-G. Barton and coworkers found that conformational gating to be highly regulating in DNA hole transport (HT), with disruptions of the DNA π stack generally yielding slower HT rates [49, 55]. For simplicity, we assume that SE transitions across mismatches, bulges, and possibly 8-oxo-G, are feasible and comparable to SE transitions through a well-matched A/T bridge. Note that the validity of this assumption is significant for model accuracy. If SE transitions across bulges are not feasible or very slow compared to thermally induced hopping (TIH) rates through bulges, the resulting CT rate across the full duplex strand becomes increasingly dependent on the (currently rough) estimate of the TIH rate through bulges, which are expected to yield significantly slower full duplex CT rates when it involves bulges or mismatches. If SE across a bulge turns out not to be possible, one could probably utilize co-factors with the 8-oxo-G located just below the ligation site and still support co-factor replication. See Section 3 for details.

In our simulations, for simplicity we assume charges are contained in DNA, i.e., there is no charge transport between DNA strands and the solvent. Note that solvent interactions

have been argued to be a significant factor in CT through mismatched DNA duplexes [62]. This is expected to decrease the CT rates in DNA duplex due to the partial loss of charges to the solvent.

Finally, we have conducted our CT simulations by assuming the HT starts from the first base pair up to the oxoG location. More correctly, our CT simulation should have started between the intercalated Ru-C and the nearest guanines within the DNA duplex. This simplification probably underestimates CT rates.

About (iii): See Section 4. Reaction kinetic seems to be the natural choice for investigating the LIDA based replication dynamics for multiple (≥ 100) strands. We have made several simplification for these kinetic LIDA simulations.

In our simulations we assume the involved DNA template strands and precursor oligomers interact in bulk solution. They are in fact vesicle-anchored DNA duplex strands, where the supplied oligomer resources consist of two oligomers with and two oligomers without membrane anchors. More realistic kinetic rates could therefore be obtained if the relevant rates were adjusted to fit the DNA anchoring. Preliminary simulation investigations of such a system indicate that the relative replication rates differences between different sequences are conserved, although the overall replication rate is expected to be faster mainly due to a faster collision rates in 2D compared to 3D [63].

Further, our LIDA kinetic equations do not include pseudo-blunt-end ligation, which is a process by which non-template-directed ligation is known to occur in the LIDA system [43]. The inclusion of pseudo-blunt-end ligation becomes important for predicting simulation dynamics of LIDA when the initial template concentrations are extremely low, as amplification can be observed in the physical system even when the initial template concentration is zero. However, the kinetic rate of blunt-end ligation is estimated to be orders of magnitude smaller than the rates of template-directed ligation. Thus, pseudo-blunt-end ligation can therefore be assumed to have only minor influence on the dynamics of LIDA when there is an initial template concentration of a few nM.

Finally, we assume a constant and sequence-independent on-rate, which simplifies the calculations, although we are aware that this is always true. For example, sequences with many base pairs repeats are known to initially mismatch followed by slower sliding and re-hybridization processes that eventually leads to correct hybridization state. These sliding and re-hybridization processes slow down the resulting hybridization process. However, assuming constant on-rates, we only need to estimate the off-rates based on the fundamental relation between the free energy of DNA hybridization and the equilibrium constant, which is equivalent to the fraction between the on-rate and the off-rate. As the off-rates are orders of magnitude slower than the on-rates, an assumption about constant on-rates will presumably

not significantly influence the results.

About (iv): See Section 5. As already previously emphasized, no interaction dynamics between the charge transport and the replication dynamics nor with the environment are taken into consideration. Thus, a combination of the simulation results found under (ii) and (iii) is a simplified (superposition), but still interesting, representation of what a more complex simulation would provide.

Given the above assumptions, we observe that viable protocells require minimum thresholds for the reaction rate constants for both co-factor replication and charge transport. Thus, protocells cannot exist with production rates below a certain threshold. These threshold rates are estimated by requiring that the overall production rate, an aggregation of the replication and the charge transfer rates, are higher than the involved degradation rates, see Sections 5 and 5.1.1.

To assess how likely it is to pick a viable co-factor at random from the available ensemble of the 160 investigated strands, we use functional information as defined by [41]. Estimation of the functional information is generally difficult, but in our situation we can make functional information estimates based on the rate estimates in Section 3 and 4. For details see Section 5.1.2.

7 Conclusion

The main result of our investigation emerge as we compare the estimated k_{CT} and k_{rep} values with the associated observed degradation rates, which indicate that protocellular viability cannot be rejected for 63 of the 160 tested co-factors, given the assumptions that underpin our estimations.

The work is based on an operational definition of minimum life that can be supported by four interconnected functionalities: a metabolic energy transducer, an informational co-factor, and a container, which must all be situated in an appropriate environment, see [1, 27, 36]. We investigate the molecular requirements of a co-factor-modulated energy transducer, assuming the co-factor is a short DNA duplex containing an 8-oxo-guanine that is able to replicate and where the energy transducer is a ruthenium complex. We present simulation results for charge transfer (CT) and replication abilities for 160 DNA co-factors each composed of 18 base pairs [1, 27, 33–36].

CT simulations are performed for DNA duplexes using charge transfer kinetics, based on an ODE model where the transition rate constants k_{CT} are determined from both theoretical calculations and empirical studies in the literature. A few CT rates had to be guesstimated by the authors (e.g., charge transfer across a bulge). Typical CT rate constants k_{CT} range

from $10^{-1}/\text{s}$ to $10^6/\text{s}$ for the 160 investigated DNA duplex strands, where the resulting k_{CT} is estimated from the first base of the 18 bp long strand to an 8-oxo-G located somewhere along the duplex strand.

The observed metabolic fatty acid production rate in a system with a direct ruthenium to 8-oxo-G charge transfer (without an intermediate DNA charge transfer) is measured to be about $1.3 \times 10^6/\text{s}$ and a corresponding fatty acid (and vesicle) doubling time of about 4.3 hours [33] [34] [38] [57]. If we require the DNA CT to slow down this rate no more than 1%, we require that $k_{CT} > 50/\text{s}$.

Replication dynamics of the combinatorial DNA co-factor is estimated by lesion-induced DNA amplification (LIDA) simulations using reaction kinetics ODEs, where reaction constants are determined directly from thermodynamic calculations and observed non-enzymatic ligation rate constant [35]. The obtained overall replication constants k_{rep} typically range from just below $10^{-4}/\text{s}$ to above $10^{-3}/\text{s}$, which means co-factor doubling times from below 20 minutes to many hours. Oligomer degradation rate constants are observed to be up to about $10^{-5}/\text{s}$ (imidazole hydrolysis), so we require that $k_{rep} > 10^{-4}/\text{s}$ to ensure a viable co-factor replication.

The protocellular viability is further quantified using the concept of functional information, where the needed minimal information in a co-factor is estimated. Assuming a co-factor duplex is randomly selected from the sample of 160 investigated duplexes 1.26 bits are required to satisfy the charge transport rate constant requirement, 0.89 bits are required to satisfy the replication rate requirement, while both requirements need 1.34 bits of information.

Acknowledgements

Part of the computation done for this project was performed on the UCloud interactive HPC system, which is managed by the eScience Center at the University of Southern Denmark. KRT and SR were in part supported by the Donostia International Physics Center (DIPC), San Sebastian, Spain. AK was partly supported by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101068029, and by Grant 62828 from the John Templeton Foundation. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation.

References

- [1] Steen Rasmussen, Mark A. Bedau, Liaohai Chen, David Deamer, David C. Krakauer, Norman H. Packard, and Peter F. Stadler. *Protocells: Briding nonliving and living matter*. MIT Press, 2008, new print 2022.
- [2] Katarzyna P Adamala, Aaron E Engelhart, and Jack W Szostak. Collaboration between primitive cell membranes and soluble catalysts. *Nature communications*, 7(1):11041, 2016.
- [3] Taro Furubayashi, Tadashi Nakano, Andrew Eckford, Yutaka Okaie, and Tetsuya Yomo. Packet fragmentation and reassembly in molecular communication. *IEEE transactions on nanobioscience*, 15(3):284–288, 2016.
- [4] Philip J Kitson, Stefan Glatzel, Wei Chen, Chang-Gen Lin, Yu-Fei Song, and Leroy Cronin. 3d printing of versatile reactionware for chemical synthesis. *Nature protocols*, 11(5):920–936, 2016.
- [5] Yan Qiao, Mei Li, Richard Booth, and Stephen Mann. Predatory behaviour in synthetic protocell communities. *Nature chemistry*, 9(2):110–119, 2017.
- [6] Laura Rodríguez-Arco, Mei Li, and Stephen Mann. Phagocytosis-inspired behaviour in synthetic protocell communities of compartmentalized colloidal objects. *Nature materials*, 16(8):857–863, 2017.
- [7] Masayuki Imai, Yuka Sakuma, Minoru Kurisu, and Peter Walde. From vesicles toward protocells and minimal cells. *Soft Matter*, 18(26):4823–4849, 2022.
- [8] Wakana Sato, Melanie Rasmussen, Christopher Deich, Aaron E Engelhart, and Katarzyna P Adamala. Expanding luciferase reporter systems for cell-free protein expression. *Scientific reports*, 12(1):11489, 2022.
- [9] Petra Schwille, Joachim Spatz, Katharina Landfester, Eberhard Bodenschatz, Stephan Herminghaus, Victor Sourjik, Tobias J Erb, Philippe Bastiaens, Reinhard Lipowsky, Anthony Hyman, et al. Maxsynbio: avenues towards creating cells from the bottom up. *Angewandte Chemie International Edition*, 57(41):13382–13392, 2018.
- [10] Kevin Jahnke, Marian Weiss, Christoph Frey, Silvia Antona, Jan-Willi Janiesch, Ilia Platzman, Kerstin Göpfrich, and Joachim P Spatz. Programmable functionalization of surfactant-stabilized microfluidic droplets via DNA-tags. *Advanced Functional Materials*, 29(23):1808647, 2019.

- [11] Hub Zwart. From primal scenes to synthetic cells. *Elife*, 8:e46518, 2019.
- [12] Tjeerd Pols, Hendrik R Sikkema, Bauke F Gaastra, Jacopo Frallicciardi, Wojciech M Śmigiel, Shubham Singh, and Bert Poolman. A synthetic metabolic network for physicochemical homeostasis. *Nature Communications*, 10(1):4239, 2019.
- [13] John I Glass, Nacyra Assad-Garcia, Nina Alperovich, Shibu Yooseph, Matthew R Lewis, Mahir Maruf, Clyde A Hutchison III, Hamilton O Smith, and J Craig Venter. Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences*, 103(2):425–430, 2006.
- [14] Marco Villani, Alessandro Filisetti, Alex Graudenzi, Chiara Damiani, Timoteo Carletti, and Roberto Serra. Growth and division in a dynamic protocell model. *Life*, 4(4):837–864, 2014.
- [15] Ben Shirt-Ediss, Ricard V Solé, and Kepa Ruiz-Mirazo. Emergent chemical behavior in variable-volume protocells. *Life*, 5(1):181–211, 2015.
- [16] J Doyne Farmer, Stuart A Kauffman, and Norman H Packard. Autocatalytic replication of polymers. *Physica D: Nonlinear Phenomena*, 22(1-3):50–67, 1986.
- [17] Vera Vasas, Chrisantha Fernando, Mauro Santos, Stuart Kauffman, and Eörs Szathmáry. Evolution before genes. *Biology direct*, 7:1–14, 2012.
- [18] Shinpei Tanaka, Harold Fellermann, and Steen Rasmussen. Structure and selection in an autocatalytic binary polymer model. *Europhysics Letters*, 107(2):28004, 2014.
- [19] Katarzyna Adamala and Jack W Szostak. Nonenzymatic template-directed RNA synthesis inside model protocells. *Science*, 342(6162):1098–1100, 2013.
- [20] Dirk Sievers and Günter von Kiedrowski. Self-replication of hexadeoxynucleotide analogues: autocatalysis versus cross-catalysis. *Chemistry—A European Journal*, 4(4):629–641, 1998.
- [21] Tracey A Lincoln and Gerald F Joyce. Self-sustained replication of an RNA enzyme. *Science*, 323(5918):1229–1232, 2009.
- [22] Harold Fellermann, Bernat Corominas-Murtra, Per Lyngs Hansen, John Hjort Ipsen, Ricard Solé, and Steen Rasmussen. Non-equilibrium thermodynamics of self-replicating protocells. 2015.

- [23] Harold Fellermann, Omer Markovitch, Owen Gilfellon, Curtis Madsen, and Andrew Phillips. Toward programmable biology. *ACS Synthetic Biology*, 5(8):793–794, 2016.
- [24] Bernat Corominas-Murtra. Thermodynamics of duplication thresholds in synthetic protocell systems. *Life*, 9(1):9, 2019.
- [25] Jeremy L England. Statistical physics of self-replication. *The Journal of chemical physics*, 139(12):09B623_1, 2013.
- [26] Artemy Kolchinsky. Thermodynamic dissipation does not bound replicator growth and decay rates. *arXiv preprint arXiv:2404.01130*, 2024.
- [27] Steen Rasmussen, Liaohai Chen, Martin Nilsson, and Shigeaki Abe. Bridging nonliving and living matter. *Artificial life*, 9(3):269–316, 2003.
- [28] Timoteo Carletti, Roberto Serra, Irene Poli, Marco Villani, and Alessandro Filisetti. Sufficient conditions for emergent synchronization in protocell models. *Journal of theoretical biology*, 254(4):741–751, 2008.
- [29] Tristan Rocheleau, Steen Rasmussen, Peter E Nielsen, Martin N Jacobi, and Hans Ziock. Emergence of protocellular growth laws. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1486):1841–1845, 2007.
- [30] Steen Rasmussen, Mark A. Bedau, and Norman H. Packard John S. McCaskill. A roadmap to protocells. *Protocells: Bridging nonliving and living matter*, pages 71 – 100, 2018, new print 2022.
- [31] S Rasmussen, L Chen, D Deamer, DC Krakauer, NH Packard, PF Stadler, and MA Bedau. Transitions from nonliving to living matter. *Science*, 303:963–965, 2004.
- [32] Harold Fellermann, Steen Rasmussen, Hans J Ziock, and Ricard V Solé. Life cycle of a minimal protocell – a dissipative particle dynamics study. *Artificial Life 13*, 13(4):319–345, 2007.
- [33] Michael S DeClue, Pierre-Alain Monnard, James A Bailey, Sarah E Maurer, Gavin E Collis, Hans-Joachim Ziock, Steen Rasmussen, and James M Boncella. Nucleobase mediated, photocatalytic vesicle formation from an ester precursor. *Journal of the American Chemical Society*, 131(3):931–933, 2009.

- [34] Sarah E Maurer, Michael S DeClue, Anders N Albertsen, Mark Dörr, David S Kuiper, Hans Ziock, Steen Rasmussen, James M Boncella, and Pierre-Alain Monnard. Interactions between catalysts and amphiphilic structures and their implications for a protocell model. *ChemPhysChem*, 12(4):828–835, 2011.
- [35] Jonathan L Cape, Joseph B Edson, Liam P Spencer, Michael S DeClue, Hans-Joachim Ziock, Sarah Maurer, Steen Rasmussen, Pierre-Alain Monnard, and James M Boncella. Phototriggered DNA phosphoramidate ligation in a tandem 5'-amine deprotection 3'-imidazole activated phosphate coupling reaction. *Bioconjugate Chemistry*, 23(10):2014–2019, 2012.
- [36] Steen Rasmussen, Adi Constantinescu, and Carsten Svaneborg. Generating minimal living systems from non-living materials and increasing their evolutionary abilities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1701):20150440, 2016.
- [37] Jan Engelhardt, Bent Andresen, and Steen Rasmussen. Thermodynamics and kinetics of Lesion Induced DNA Amplification (LIDA). In *ALIFE 2020: The 2020 Conference on Artificial Life*, pages 269–272. MIT Press, 2020.
- [38] Ditlev Hartmann Bornebusch, Hans-Joachim Ziock, and Steen Rasmussen. Reaction kinetics exploration of a protocellular metabolism. *ALIFE 2021: The 2021 Conference on Artificial Life*, 07 2021.
- [39] Kristoffer R. Thomsen. *Energy, information and the emergence of living matter*. University of Southern Denmark, 2022.
- [40] Jack W Szostak. Functional information: Molecular messages. *Nature*, 423(6941):689–689, 2003.
- [41] Robert M Hazen, Patrick L Griffin, James M Carothers, and Jack W Szostak. Functional information and the emergence of biocomplexity. *Proceedings of the National Academy of Sciences*, 104(suppl 1):8574–8581, 2007.
- [42] G von Kiedrowski. Minimal replicator theory I: Parabolic versus exponential growth. *Bioorganic Chemistry Frontiers*, 3, 1993.
- [43] B Safeenaz Alladin-Mustan, Catherine J Mitran, and Julianne M Gibbs-Davis. Achieving room temperature DNA amplification by dialling in destabilization. *Chemical Communications*, 51(44):9101–9104, 2015.

- [44] Alexander A Voityuk. Estimates of electronic coupling for excess electron transfer in DNA. *The Journal of chemical physics*, 123(3):034903, 2005.
- [45] Steffen Loft and Henrik E Poulsen. Cancer risk and oxidative DNA damage in man. *Journal of molecular medicine*, 74(6):297–312, 1996.
- [46] Krzysztof Roszkowski, Wojciech Jozwicki, Piotr Blaszczyk, Anna Mucha-Malecka, and Agnieszka Siomek. Oxidative damage DNA: 8-oxogua and 8-oxodg as molecular markers of cancer. *Medical science monitor: international medical journal of experimental and clinical research*, 17(6):CR329, 2011.
- [47] T Gregory Drummond, Michael G Hill, and Jacqueline K Barton. Electrochemical DNA sensors. *Nature biotechnology*, 21(10):1192–1199, 2003.
- [48] Aleksandra Mihailovic, Ioana Vladescu, Micah McCauley, Elaine Ly, Mark C Williams, Eileen M Spain, and Megan E Nuñez. Exploring the interaction of ruthenium (ii) polypyridyl complexes with DNA using single-molecule techniques. *Langmuir*, 22(10):4699–4709, 2006.
- [49] Joseph C Genereux and Jacqueline K Barton. Mechanisms for DNA charge transport. *Chemical reviews*, 110(3):1642–1662, 2010.
- [50] Tashica T Williams, Duncan T Odom, and Jacqueline K Barton. Variations in DNA charge transport with nucleotide composition and sequence. *Journal of the American Chemical Society*, 122(37):9048–9049, 2000.
- [51] Susan M Gasper and Gary B Schuster. Intramolecular photoinduced electron transfer to anthraquinones linked to duplex DNA: the effect of gaps and traps on long-range radical cation migration. *Journal of the American Chemical Society*, 119(52):12762–12771, 1997.
- [52] M Bixon and Joshua Jortner. Long-range and very long-range charge transport in DNA. *Chemical Physics*, 281(2-3):393–408, 2002.
- [53] Yasuko Osakada, Kiyohiko Kawai, Mamoru Fujitsuka, and Tetsuro Majima. Kinetics of charge transfer in DNA containing a mismatch. *Nucleic acids research*, 36(17):5562–5570, 2008.
- [54] John SantaLucia Jr and Donald Hicks. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33:415–440, 2004.

- [55] Pratip K Bhattacharya and Jacqueline K Barton. Influence of intervening mismatches on long-range guanine oxidation in DNA duplexes. *Journal of the American Chemical Society*, 123(36):8649–8656, 2001.
- [56] Nicholas F Polizzi, Michael J Therien, and David N Beratan. Mean first-passage times in biology. *Israel journal of chemistry*, 56(9-10):816–824, 2016.
- [57] Ditlev Hartmann Bornebusch. Reaction kinetic exploration of a protocellular metabolism and information. Master’s thesis, University of Southern Denmark, 2021.
- [58] Kunio Kawamura and Jun Maeda. Kinetic analysis of oligo(c) formation from the 5'-monophosphorimidazolide of cytidine with pb(ii) ion catalyst at 10-75°C. *Orig Life Evol Biosph*, (37):153–165, 2007.
- [59] Ulla Christensen, Nana Jacobsen, Vivek K Rajwanshi, Jesper Wengel, and Troels Koch. Stopped-flow kinetics of locked nucleic acid (LNA)oligonucleotide duplex formation: studies of LNA-DNA and DNA-DNA interactions. (354):481–484, 2001.
- [60] Hatim T Allawi and John SantaLucia. Thermodynamics and NMR of internal G·T mismatches in DNA. *Biochemistry*, 36(34):10581–10594, 1997.
- [61] Hatim T Allawi and John SantaLucia. Nearest neighbor thermodynamic parameters for internal g·a mismatches in DNA. *Biochemistry*, 37(8):2170–2179, 1998.
- [62] Bernd Giese and Stefan Wessely. The influence of mismatches on long-distance charge transport through DNA. *Angewandte Chemie International Edition*, 39(19):3490–3491, 2000.
- [63] Marco Tuccio. Zipping graph model of DNA hybridization kinetics, predictions on the membrane anchored case and application to lesion induced DNA amplifications systems, 2023.