PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Reinholt Thomsen K, Kolchinsky A, Rasmussen S. 2025 Protocellular energy transduction, information and fitness. *Phil. Trans. R. Soc. B* **380**: 20240294.
https://doi.org/10.1098/rstb.2024.0294

Received: 1 October 2024
Accepted: 22 April 2025

One contribution of 15 to a theme issue 'Origins of life: the possible and the actual'.

Subject Areas:
biomaterials, biophysics, computational biology, evolution, synthetic biology

Keywords:
protocells, origins of life, metabolic networks, charge transport, replication, fitness

Author for correspondence:
Steen Rasmussen
e-mail: steen@sdu.dk

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.8051168.

THE ROYAL SOCIETY PUBLISHING

Protocellular energy transduction, information and fitness

Kristoffer Reinholt Thomsen[1], Artemy Kolchinsky[2,3] and Steen Rasmussen[1,3,4]

[1]Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, 5230 Odense, Syddanmark, Denmark
[2]Department of Information, Indiana University Bloomington School of Informatics and Computing, Bloomington, IN 47408, USA
[3]Santa Fe Institute, Santa Fe, NM 87501, USA
[4]Universita Ca' Foscari European Centre for Living Technology, 30123 Venice, Italy

AK, 0000-0002-3518-9208; SR, 0000-0002-3336-843X

An integrated protocellular system is presented and studied in simulation. In this system, co-factor molecules serve both as part of an electron relay for an energy transduction mechanism and as combinatorial molecules that can replicate. Thus, the co-factor acts as a primitive inheritable information store. This work is done to elucidate critical experimental design issues connecting energy transduction and inheritable information. Simulations are used to study how the co-factor sequence determines the protocellular 'fitness' as reflected by the calculated charge transfer and replication rates, and we compare these rates with experimentally observed production and degradation rates from similar systems. Replication and charge transport processes turn out to have different and often opposing co-factor requirements, and it turns out that we can estimate the combined fitness of the two processes. Finally, we probe the feasibility of randomly picking co-factor molecules from a limited population, where a good co-factor can enhance both metabolic biomass production and its own replication rate.

This article is part of the theme issue 'Origins of life: the possible and the actual'.

1. Introduction

In all bottom-up artificial cell 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 physicochemical systems is crucial. 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 kept together by a container.

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*. [2], Yomo *et al*. [3], Cronin *et al*. [4], Mann *et al*. [5,6], Rasmussen *et al*. [7], Adamala *et al*. [8], the Schwille *et al*. and

the MaxSynBio consortium [9,10], and perhaps the most ambitious, the ongoing Dutch Synthetic Cell project [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 container growth and division [14,15], container functionalities [8], formation of autocatalytic metabolic sets [16–18], information replication [19–22], thermodynamics of replication [23–27], protocell integration [7,19,28], inter-protocell interactions [5,6], as well as coupling and synchronization of growth among the different protocell processes [29,30]. A comparative discussion of many of these different approaches can be found in Rasmussen *et al*. [31].

This article 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 that also acts as an information carrier [28,30,32–40].

In §2, we present and discuss the key features of the proposed protocell design, as well as the functionalities essential for the following analysis. In §§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 article. We perform 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 [37]. The simulations in §§3 and 4 may therefore be seen as preparations for future experiments. In §5, we compare and contrast the results from §§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 [41,42] to our findings to get a sense of how likely it is to pick a 'good' co-factor at random from a limited polymer sequence ensemble, and in §6 we draw the conclusions from our study. Electronic supplementary material expands multiple details about the reported work, including critical discussions of the underlying model assumptions as well as the main findings.

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 through autocatalysis [28]. 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 two-dimensional containers; see figure 1 (1) where four coupled energy transducers and co-factors are depiced at the exterior of a vesicle surface. (For more details, see electronic supplementary material, figure S1)

The combined energy transducer and co-factor, 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 made up of decanoic acid building blocks 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 discussion of each of the involved processes is found in the electronic supplementary material with illustrations figures S2–S4. A summary discussion of the main processes in the protocellular life-cycle is shown in figure 1.

DNA is used as a combinatorial co-factor in the metabolism because of its well-known replication property, and because the DNA base composition and sequence critically impact its charge transfer properties and thus the metabolic efficiency. A protocellular inheritance is obtained by replication of the co-factor. A detailed analysis of the DNA charge (electron hole) transfer properties is the topic of §3. Due to the well-known product inhibition in template-directed ligation replication [43], we employ DNA sequences that are capable of self-replication via isothermal lesion-induced DNA amplification (LIDA) [44]. A detailed analysis of the co-factor replication properties is the topic of §4.

Both DNA charge transfer and non-enzymatic DNA replication properties strongly depend on the DNA sequence composition. As we shall see in the following sections, the charge transfer and replication properties have very different sequence requirements, which makes it nontrivial to identify appropriate sequences that can support both functionalities. It should be noted that the functional information stored in the co-factor is neither purely sequential nor purely compositional [22] in nature. As we shall see in §§3, 4 and 5, the co-factor information is determined by its ability to act both as a charge transfer device and a replicator. These properties depend on which bases are present (compositional information) as well as on their position relative to the membrane (sequence-based information). Furthermore, it should be noted that the environment (e.g. temperature, pH, salts) must be highly regulated 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 and protection groups), Ru complexes with tails and dihydrophenyl glycine (hydrogen source).

3. Simulation exploration of co-factor charge transport

As the overall protocellular metabolic process, by design, is controlled by charge transport via a double-stranded DNA (dsDNA), we explore how the sequence and composition details of the DNA impact its charge transfer capabilities.

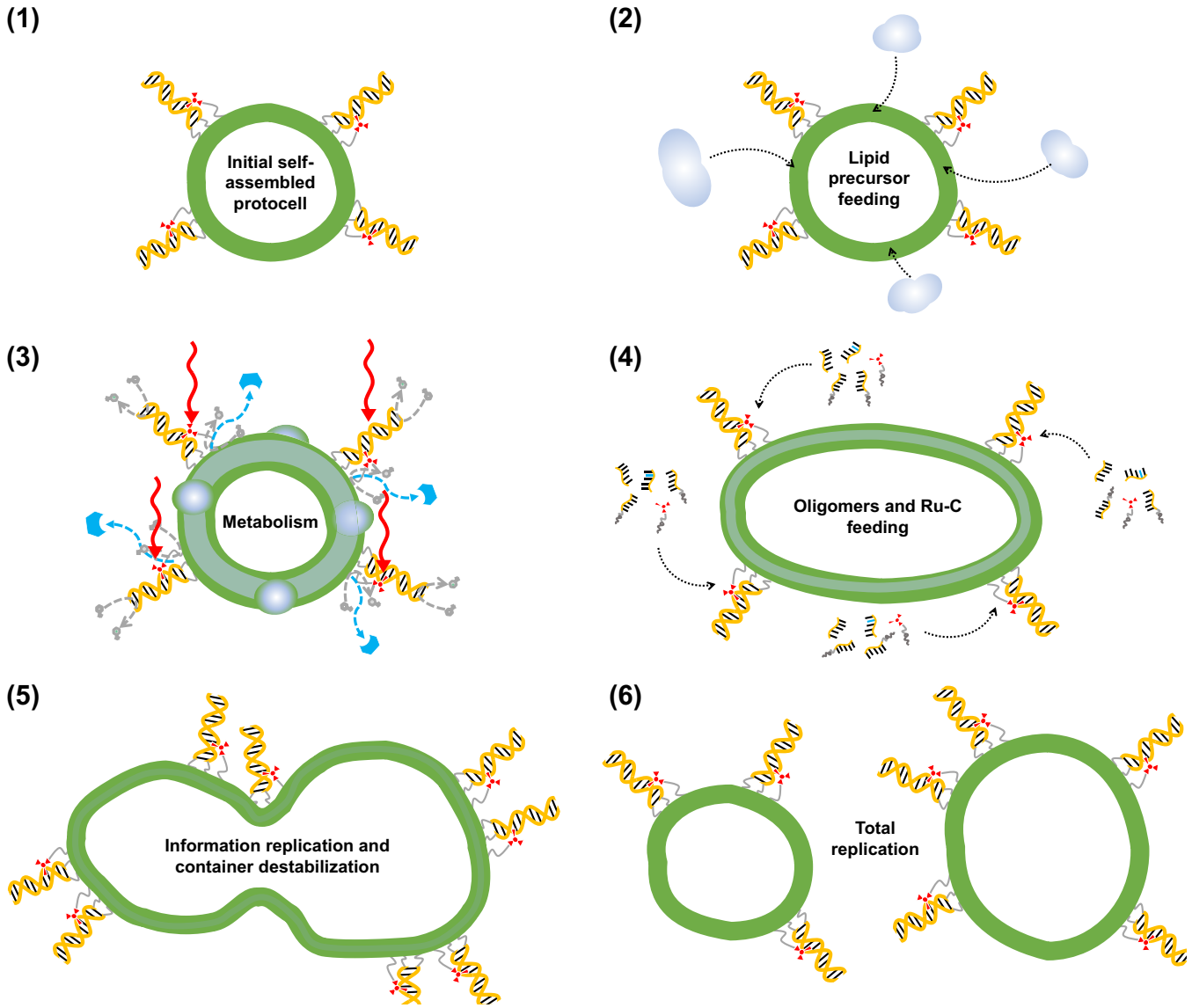


Figure 1. Protocellular life-cycle: (1) A cartoon of the protocell with a fatty (decanoic) acid vesicle container decorated with anchored ruthenium complexes and DNA duplexes. (2) Droplet feeding of hydrophobic membrane precursors (picolineum esters). (3) Absorption of the hydrophobic droplets into the membrane, where they are partly dissolved in the membrane, while the metabolism converts precursor lipids (picolinium ester) into lipids (decanoic acid) through photo-deprotection. This part of the metabolic process continues through (4). Because this metabolic process is modulated by DNA co-factor charge transfer, the rate of overall resource conversion depends on the charge transfer rate constant k_{CT}, discussed in §3. (4) Feeding of ruthenium complexes and DNA oligomers. Note that all precursors can be fed at once but are shown here as a two-step process for clarity. The rate of replication of the DNA co-factor is specified by the rate constant k_{rep}, discussed in §4. (5) A completed (DNA) co-factor replication and fatty acid production, which together result in membrane growth and eventually vesicle destabilization. The original vesicle eventually breaks up and forms two new protocells, (6). The overall protocellular fitness (rate of protocell growth and division) depends on a nontrivial, quantitative combination of k_{CT} and k_{rep}, discussed in §5. A more detailed discussion of each reaction step (1)–(6) is found in electronic supplementary material, §A.

Charge transport (CT) in dsDNA has been extensively studied since the early 1990s. The majority of DNA CT research has focused on hole transport (HT), with less research devoted to electron transport (ET) [45]. 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 [46,47]. Therefore, electrochemistry-based sensors have been proposed for the detection of selected DNA sequences or mutated genes associated with human disease [48].

Our previous explorations of protocellular metabolic function have focused on direct ET from oxoG to a Ru-C, either covalently bound or in close spactial proximity [34–37,39]. 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 involves both 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) the missing electron in the guanine (the hole) then travels through the DNA strand via HT, until it meets an oxoG that donates the missing electron. This is illustrated in figure 2 left panel. To repair the now damaged oxoG, a proton donor (dihydrophenylglycine) is provided to the system, as in our previous studies.

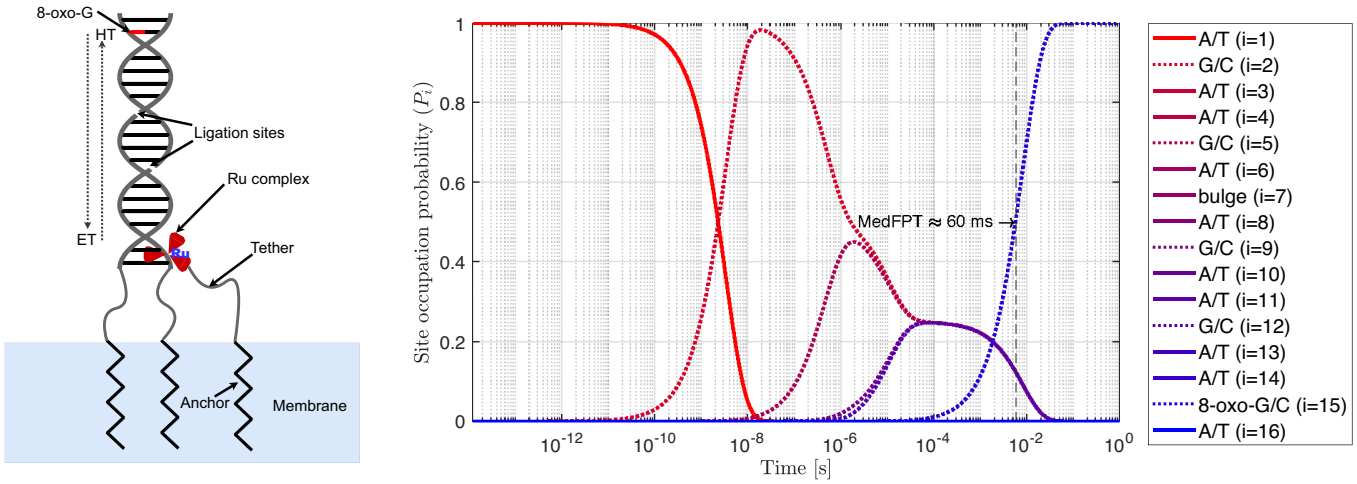


Figure 2. *Left:* Example of protocellular metabolism design. The co-factor containing oxoG and Ru-C are anchored to the protocellular container surface via tethers to amphiphilic anchors. We assume the Ru-C intercalates with the DNA π stack of the co-factor two bases from the 'bottom', enabling CT between Ru-C and 8-oxo-G mediated by the nucleobases of the DNA strand, see text for details. A fast DNA HT (= CT) means a faster metabolism, while a slow HT means a slower metabolism. Examples of DNA oligomer ligation sites are shown, since they impact the HT rates. *Right:* CT simulation results for strand number 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 hole occupation probabilities for A/T base pairs are very low compared with base pairs containing G or 8-oxo-G.

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 [49–51]. It is also known that HT occurs in DNA strands containing oxoG, and that oxoG is a deep thermodynamic potential well for holes [52]. 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 the left panel of figure 2. Based on these experimental facts, we propose a protocellular design in which DNA strands, containing a single oxoG near one terminal, are attached to the protocell container (membrane) via hydrophobic anchor at the opposite terminal. Also, Ru-C (energy transducers) are anchored to the membrane via molecular tethers, thereby promoting intercalation with the π stack of the DNA duplex.

(a) Modelling charge transport reactions

Because the rate of the DNA CT is critical for the metabolism, we explore how the CT is influenced by the composition of the DNA strand. We model DNA CT as a system of N redox centres whose occupation is governed by the following master equation:

$$\frac{\mathrm{d}P_i(t)}{\mathrm{d}t} = \sum_{j=1}^{N} \kappa_{ij} P_j(t) \quad \text{where} \quad i \in \{1, .., N\}. \tag{3.1}$$

$P_i(t)$ indicates the probability that the charge is located at the ith redox centre at time t, and κ_{ij} are elements of a matrix containing the transition rates between redox centres (κ_{ij} is the transition rate to the ith from the jth redox centre). Assuming there is no loss of charge to the environment, the rate matrix should conserve charge, $\sum_i \kappa_{ij} = 0$.

We may now use the rate matrix formalism to model CT in DNA. Because the transition rates depend on quantum tunnelling rates, they decay exponentially as a function of the distance between redox centres. In models that only account for Thermally Induced Hole transport transitions (TIH), transition rates far from the diagonal of the κ_{ij} matrix are set to zero, yielding a band matrix. If transitions to non-neighbouring bases are allowed—e.g. by including the super exchange (SE) mechanism [53,54]—the rate matrix will no longer be a simple band matrix, but instead have non-zero entries away from the band. This is the case for the example rate matrix in equation (A2) in the electronic supplementary material that represents a simple well-matched strand 5′-AGAAGA-3′:

$$\boldsymbol{\kappa} = \begin{bmatrix} -k_{AG} & k_{GA} & 0 & 0 & 0 & 0 \\ k_{AG} & -(2k_{GA} + k_{GG}(2)) & k_{AG} & 0 & k_{GG}(2) & 0 \\ 0 & k_{GA} & -(k_{AG} + k_{AA}) & k_{AA} & 0 & 0 \\ 0 & 0 & k_{AA} & -(k_{AA} + k_{AG}) & k_{GA} & 0 \\ 0 & k_{GG}(2) & 0 & k_{AG} & -(2k_{GA} + k_{GG}(2)) & k_{AG} \\ 0 & 0 & 0 & 0 & k_{GA} & -k_{AG} \end{bmatrix} \tag{3.2}$$

Table 1. Summary of CT rates in our simulations.

name	rate (s^{-1})	reference	description
k_{AA}	5×10^7	Bixon & Jortner [53]	rate from A/T to A/T
k_{GA}	4×10^4	Bixon & Jortner [53]	rate from G/C to A/T
k_{AG}	3×10^8	Bixon & Jortner [53]	rate from A/T to G/C
$k_{GG}(n=1)$	1.2×10^8	Bixon & Jortner [53][a]	super-exchange rate from G/C to G/C
k_{AoxoG}	4×10^6	estimation by authors	rate from A/T to oxoG/C
k_{oxoGA}	0	estimation by authors[b]	rate from oxoG/C to A/T
k_{GG_T}	1×10^4	Osakada *et al*. [55]	rate through substrand from G/C to G/T
k_{GG_A}	9.5×10^3	Osakada *et al*. [55]	rate through substrand from G/C to G/A
k_{AB}	1×10^3	estimation by authors	rate from A/T to bulge
k_{BA}	1×10^3	estimation by authors	rate from bulge to A/T

[a]The 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 is determined by the electronic couplings in the bridge or alternatively a β -value specific for the bridge ($r = \exp(-\beta R_0)$), where R_0 is the base separation in the DNA stack. In our investigations, we use $\beta = 0.7$ A^{-1} and $R_0 = 3.4$ A. In our simulations, we set a maximum bridge length of $n = 9$ bp.
[b]Since oxoG is a thermodynamic trap, the rate out of oxoG is assumed to be very low compared with the rest of the used rates. K_{oxoGA} is therefore set equal to zero for simplicity.

Table 2. Substrands used to generate co-factor sequences for the metabolism and replication simulations. Substrand 1 and 2 are well matched while substrand 3 and 4 are mismatched. Substrand 5 has substituted a guanine for an 8-oxoguanine, and substrand 6 contains a bulging adenine.

substrand 1	substrand 2	substrand 3	substrand 4	substrand 5	substrand 6
5′-AGA-3′ 3′-TCT-5′	5′-AAA-3′ 3′-TTT-5′	5′-AGA-3′ 3′-TTT-5′	5′-AGA-3′ 3′-TAT-5′	5′-A8oxoGA-3′ 3′-TCT-5′	5′-AAA-3′ 3′-TT-5′

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 neighbouring bases; (iii) holes only transfer between the bases with the lowest oxidation potential in each base pair ($E_{ox}^G < E_{ox}^C$ and $E_{ox}^A < E_{ox}^T$ [53]); (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 [53] is approximately equal to the SE rate from G to a single G[1]; (*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. In the following simulations, the charge is initially localized at the first base, i.e. $P_1(t=0) = 1$ and $P_{i\neq 1}(t=0) = 0$.

(b) Charge transport simulations in a selection of strands

To obtain physically realistic CT rates, we simulate strands with known experimental or theoretical single-step HT k_{ij} rates. For these, we use previously published results, based on either experimental data or theoretical calculations, except for the rates between (i) A/T and oxoG/C pairs, and (ii) A/T and a bulging base, which are estimated by the authors. The rates are shown in table 1.

The experimentally estimated rates are from Osakada *et al*. [55], who experimentally investigated charge transport through strands containing mismatches. They also deduced rates between certain redox sites (not necessarily neighbouring base pairs) by fitting kinetic models to their results. The theoretically estimated rates are from Bixon & Jortner [53], which are based on semi-classical Marcus theory.

As will become clear in §4 on co-factor replication, we have selected a group of 18 base pair strands containing both a bulge (an unpaired 'bulging' base) and possibly mismatches (bases forming non-Watson–Crick pairs) [56]. A bulge occurs when one base is detached from the duplex structure somewhere in the middle of the strand, e.g. if one strand of a well-matched duplex is one base longer than the other strand.

We assume that the rate through mismatches is independent of the sequence of the rest of the strand. Rates across mismatches (k_{GG_T} and k_{GG_A}) are implemented as a combined rate across multiple base pairs, as done by Osakada *et al*. [55]. Furthermore, the charge transport across the mismatch is modelled as being one-directional (no back-transfer) as by Osakada *et al*. [55]. The values of k_{GG_T} and k_{GG_A} 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 [50,57]. This motivated our estimate of HT rates through the bulge (k_{AB}

[1]This assumption is based on the energy dependent nuclear Franck–Condon factor in electronic supplementary material, equation (A2), where Bixon & Jortner [53] use $\lambda = 0.25$ eV and $\hbar\omega = 0.18$ eV, both of which are large compared with eV $\Delta G = -0.096$.

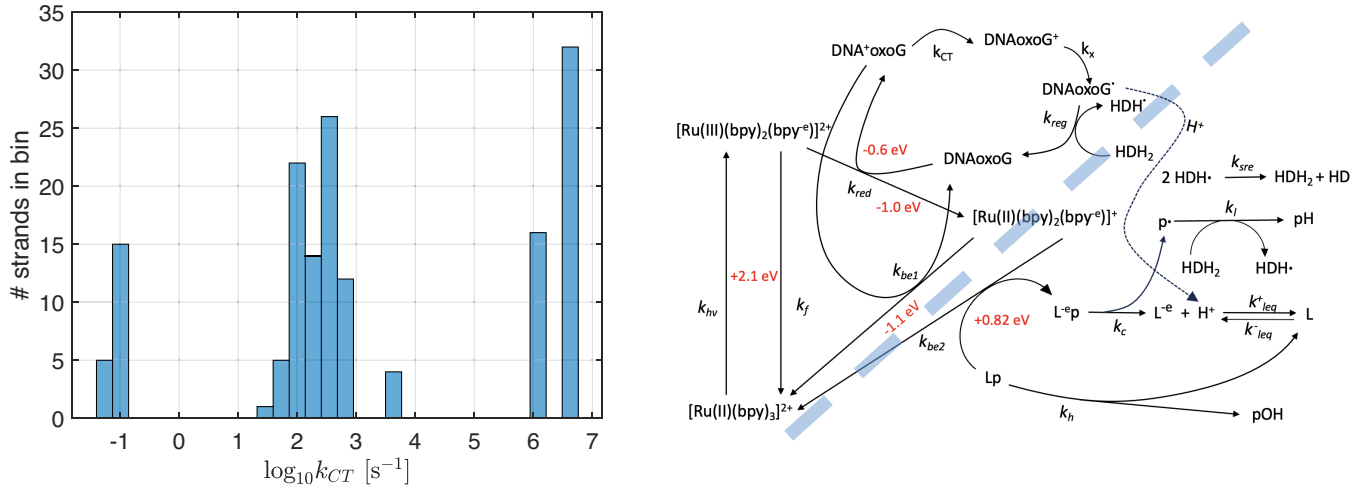


Figure 3. *Left:* Histogram showing the distribution of charge transfer rates to 8-oxo-G through the set of 160 strands. Note that CT rates cluster into three groups, see text for details. *Right:* Protocellular metabolic reaction network. See text for details. We only need to consider the upper left part of the reactions (above blue dotted line) to evaluate the overall impact of the DNA charge transfer process. (Presentation and discussion of the full network, the rate constants, and its kinetics can be found in [34,39,58]).

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 with bulges. However, we assume that SE is still possible across a bulge.

Clearly, using only the elemental HT rates listed in table 1 limits the possible strands we can simulate in a sequence of 18 bps. Table 2 shows 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_{GG_A} and k_{GG_T} to apply. Furthermore, (ii) the full strand must include one and only one bulge (substrand 6) near the centre of the strand (see §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 we consider as possible protocell co-factor candidates.

The full 160 strands are shown in the electronic supplementary material, table S2 in §D. In addition, in that table, the conclusions from §§3 and 4 are indicated by colour 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 [34–37,39], which is shown on figure 3 right panel and further discussed in §3c. For the following CT simulations, it is assumed that a Ru-C intercalates near 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 (diffuses) within the strand until it reaches oxoG, where it is trapped. A MATLAB program constructs the system of ordinary differential equations (ODEs) and solves the master equation (3.1) for each of the 160 strand sequences numerically using `ode15s`. As previously stated, the hole is assumed to be initially localized at the first site ($i = 1$), which is 2 bps up from the 'bottom' of the strand, where the anchor is attached.

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 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, the right panel in figure 2 shows the obtained temporal dynamics of CT for one of the simulated strands.

The following tendencies emerge from the 160 strand simulations: the fastest CT rates are observed when oxoG is located before the bulge, as expected. Also, SE and its distance dependence strongly influence CT rates, especially when SE occurs across the bulge. When no SE transitions are present across the bulge (when the 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 with the rate of multiple TIH jumps across the well-matched strand. Finally, the assumption of no back-transfer across mismatches yields cases where the stopping condition ($P_{8oxoG}(t) = 0.99$) 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 a mismatch, as charge is trapped by mismatches at close proximity to oxoG. The left panel of figure 3 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 $\tilde{\boldsymbol{\kappa}}$ as the rate matrix in equation (3.2) 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 $\tau_{\text{MeanFPT}} = -\sum_i [\tilde{\boldsymbol{\kappa}}^{-1}]_{i1}$ [59]. Generally, we find that MedFPT is roughly equal to $\simeq 0.7\times$ MeanFPT, except for strands 71 and 73 where MedFPT $\simeq 0.04\times$ 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.

The CT rates shown in figure 3 (left panel) show three distinct groups: (i) strands with $k_{CT} < 10^{-1}$ s^{-1}, (ii) strands with $10^0 < k_{CT} < 10^4$ and (iii) strands with $k_{CT} > 10^5$ s^{-1}. The common factor for group (i) is the absence of a SE transition across the centre 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 §2, 8-oxo-G being located before the ligation site is undesirable, since it allows metabolic processes to proceed using only the lower-resource oligomer. This bypasses the need for full replication, causing the metabolism to lose 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.

(c) Metabolic impact of DNA charge transfer

As discussed in §2, DNA charge transfer is explored as a composition and sequence-dependent process that is 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 figure 3 (right panel), in the upper left region above the blue dotted line. A detailed discussion of the full metabolic network can be found in deClue *et al*., Engelhardt and Bornebusch [34,39,58], although this earlier work did not include the double-stranded DNA charge transfer process, as it was missing the step that includes the DNAoxoG complex.

In the right panel of figure 3, the reaction constant $k_{h\nu} \simeq 0.5\ \text{s}^{-1}$ determines the photo activation rate. The fluorescent rate constant is $k_f \simeq 3 \times 10^6\ \text{s}^{-1}$/s, while the bimolecular reduction rate constant is $k_{red} \simeq 3 \times 10^5\ \text{M}^{-1}\ \text{s}^{-1}$. 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 3, right panel, 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 $k_{red} \simeq 3 \times 10^5\ \text{M}^{-1}\ \text{s}^{-1}$ and the concentration of ruthenium complex [RuC]. Next, a hole diffuses inside the DNA$^+$oxoG duplex towards the oxoguanine, whose rate constant k_{CT} we estimated for a variety of strand combination in this section and summarized in figure 3, right panel. The hole is eventually absorbed by oxoguanine (DNAoxoG$^+$), which thereby loses a hydrogen, as determined by the rate constant $k_x \simeq 10^8\ \text{s}^{-1}$ to become DNAoxoG$^\bullet$. The oxoguanine, now missing a hydrogen, is successively provided with a new hydrogen from a sacrificial hydrogen donor, HDH_2 (dihydrophenylglycine), whose rate is determined by the bimolecular rate constant $k_{reg} \simeq 10^6\ \text{M}^{-1}\ \text{s}^{-1}$ and the concentration [HDH_2]. This concludes the cycle and makes the DNA duplex ready to donate a new electron to the ruthenium complex.

The time needed to complete this circular reaction is determined by the four involved reaction constants k_{red}[RuC], k_{CT}, k_x, k_{reg}[HDH_2] and can be approximated by $t_{circular} \simeq 1/k_{red}[\text{RuC}] + 1/k_{CT} + 1/k_x + 1/k_{reg}[\text{HDH}_2]$, so the resulting reaction rate constant can be approximated as

$$\begin{aligned} k_{circular} \simeq & \{k_{red}[\text{RuC}] \times k_{CT} \times k_x \times k_{reg}[\text{HDH}_2]\}/ \\ & \{(k_{CT} \times k_x \times k_{reg}[\text{HDH}_2]) + (k_{red}[\text{RuC}] \times k_x \times k_{reg}[\text{HDH}_2]) + \\ & (k_{red}[\text{RuC}] \times k_{CT} \times k_{reg}[\text{HDH}_2]) + (k_{red}[\text{RuC}] \times k_{CT} \times k_x)\}. \end{aligned} \tag{3.3}$$

If we assume that $k_{CT} \ll k_{red}[\text{RuC}], k_x, k_{reg}[\text{HDH}_2]$, 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 [34], we have [RuC] = 1 mM and [HDH_2] = 15 mM, which means that $k_{red}[\text{RuC}] \simeq 3 \times 10^5/\text{Ms} \times 10^{-3}\text{M} \simeq 300$ /s and $k_{reg}[\text{HDH}_2] \simeq 10^6\ \text{Ms}^{-1} \times 15 \times 10^{-3}\text{M} \simeq 1.5 \times 10^3\text{s}^{-1}$. Thus, $k_{CT}(= 50\ \text{s}^{-1}) \ll k_{red}[\text{RuC}](= 300\ \text{s}^{-1})$ is only weakly fulfilled with these concentrations, as the ruthenium complex concentration is a secondary rate limiting factor for $k_{circular}$.

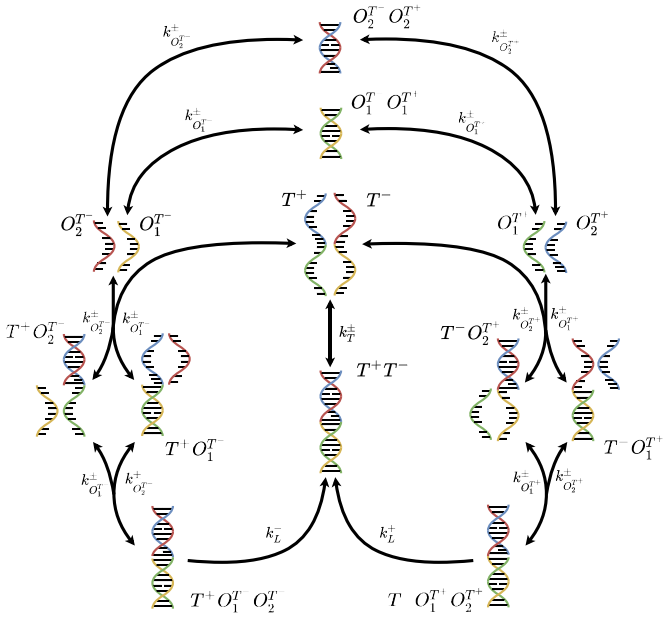
In any event, smaller k_{CT} means a smaller rate of usable photo activated electrons, as determined by

$$k_{usable-photo} \simeq \frac{k_{h\nu} \times k_{CT}}{k_{h\nu} + k_{CT}} \tag{3.4}$$

For instance, depending on whether k_{CT} is the same as $k_{h\nu}$, or is 10 times larger, or is 100 times larger than $k_{h\nu}$, the amount of usable photo activated electrons will be reduced by about 50%, 10% and 1%, respectively. In the following, we require that $k_{CT} > 100 \times k_{h\nu} = 50$/s, so that the photo activation process is still the overall rate-limiting reaction. This requirement will be used in the analysis presented in §5.

Going back to the left panel of figure 3, recall that the only strands in group (ii), with the 8-oxo-G located above the ligation site, are acceptable protocellular co-factor candidates. Furthermore, our requirement that $k_{CT} > 50\ \text{s}^{-1}$ restricts our attention to co-factors that can 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 [34,35], we measured an initial resulting fatty acid reaction rate constant of about $1.3 \times 10^{-5}\ \text{s}^{-1}$, which depends on the aggregate surface composition of picolinium ester (resource molecules) and fatty acid (products). More



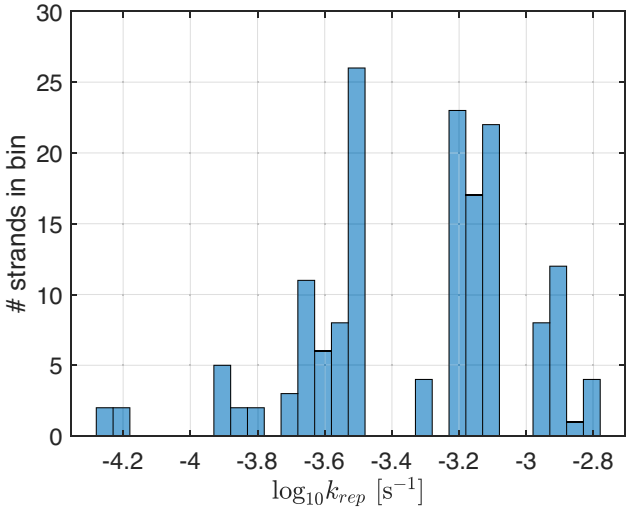


Figure 4. *Left:* The LIDA system consists of four DNA single stranded oligomers ($O_1^{T^+}$, $O_2^{T^+}$, $O_1^{T^-}$ and $O_2^{T^-}$) and two single-stranded DNA templates (T^+ and T^-). Here, the reaction scheme of LIDA with $O_1^{T^+}$, $O_2^{T^+}$, $O_1^{T^-}$ and $O_2^{T^-}$ colour coded as green, 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 $O_1^{T^\pm}$ and $O_2^{T^\pm}$ 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. *Right:* 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 $[O_1^{T^+}] = [O_1^{T^-}] = [O_2^{T^+}] = [O_2^{T^-}] = 100$ µM and $[T^-] = 14$ nM with all other initial concentrations equal to zero M. We find typical doubling times for k_{rep} between 10^{-4} s^{-1} and 10^{-3} s^{-1} where the corresponding doubling times are $\tau_{rep} = 6.70 \times 10^3$ s and 6.70×10^2 s, respectively.

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^{-1} [34], thus about an order of magnitude lower than the catalysed charge transport driven rate constant, if we 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 0.65×10^{-5} s^{-1} $\times 8.0 \times 10^{-3}$M = 5.2×10^{-7}M s^{-1}. Given this rate, τ_{double} is the time required to produce 8 mM more fatty acid from 8 mM picolinium ester concentration. Using 8×10^{-3} M = $\tau_{double} \times 5.2 \times 10^{-7}$ M s^{-1} yields $\tau_{double} = 1.54 \times 10^4$s or about 4.3 h.

4. Simulation exploration of co-factor replication

Protocellular self-replication obviously also requires replication of its DNA co-factor. Non-enzymatic molecular replication based on lesion-induced DNA amplification (LIDA) was already mentioned in §2. This section explores LIDA for the protocellular information system through simulations of kinetic equations. The full kinetic system is introduced below in the left panel of figure 4 and in the electronic supplementary material, equations (A1)–(A15) in §C [38,39,58]. 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 (electronic supplementary material, table S2 in §D) already used for the CT simulations in §3 above.

(a) The lesion-induced DNA amplification system for protocells

Alladin-Mustan *et al*. [44] reported an exponential amplification of DNA replication under isothermal conditions at room temperature by introducing destabilizing lesions into a 18 bp DNA duplex (LIDA or Lesion Induced DNA Amplification), such as abasic sites and mismatches [44]. 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.* [44] showed that the optimal temperature for DNA amplification can be tuned by the type and number of lesions added.

The total reaction scheme of the LIDA system is shown in the left panel of figure 4, but can be summarized by the reaction equation

$$O_1^- + O_2^- + O_1^+ + O_2^+ \xrightarrow{T^+ \text{ or } T^-} T^+T^- \tag{4.1}$$

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 catalysed by either simplex (T^+ or T^-) of the full DNA strand[2]. Since the duplex and simplex forms can be converted into each other via hybridization/dehybridization, the total system leads to autocatalytic production of double-stranded DNA from four single-stranded oligomers.

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

$$A + B \underset{k^-}{\overset{k^+}{\rightleftharpoons}} C \quad \text{or} \quad A \underset{k^-}{\overset{k^+}{\rightleftharpoons}} B, \tag{4.2}$$

except for the irreversible ligation reaction, where the observed rate from Cape *et al.* [36] is used. Equilibrium thermodynamics gives a relation between the standard Gibbs free energy $\Delta G°$ and the equilibrium constant K:

$$\Delta G° = -RT \ln K, \tag{4.3}$$

where $R = 1.987$ cal $\cdot$ K$^{-1}\cdot$ 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:

$$K = \frac{k^+}{k^-} = \exp\left(-\frac{\Delta G°}{RT}\right) \Rightarrow k^- = k^+ \exp\left(\frac{\Delta G°}{RT}\right). \tag{4.4}$$

Thus, if we know the on-rate (forward rate constant) k^+, the temperature T and the standard Gibbs free energy $\Delta G°$, we can compute the off-rate (reverse rate constant) k^-. For simplicity, we assume all on-rates to be $2 \cdot 10^7$ s^{-1} [60], while $\Delta G°$ is computed using the method and parameters introduced by SantaLucia & Hicks [56]. 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.

It should be noted that the 5′ end of the oligomers need to be imidazole activated for 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 [61] 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 [36], we 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 the previous §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 (A1)–(A15) of the electronic supplementary material, §C. The equations are taken from Bornebusch [58] with new additional terms, due to the possibility that 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 now used in the following §4b for simulations of the LIDA system in order to obtain the replication dynamics of protocell co-factors.

(b) Lesion-induced DNA amplification simulations for a selection of strands

By using DNA strand-specific kinetic off-rates from thermodynamic parameters and the kinetic equations of LIDA (see electronic supplementary material, §C), we simulate co-factor replication for the same 160 DNA strands that we consider in §3b when studying charge transport. This is done by splitting all strands into oligomers of equal length (9 bp or 3 substrands) where templates are composed of two oligomers. For each strand, the hybridization free energies, equation (4.3), 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 & Hicks [56]. Note that the bulge only contributes to the free energy for T^+T^-. The thermodynamic contribution of 8-oxo-G is unknown, but Gasper & Schuster [52] note that replacement of G by 8-oxo-G has little or no effect on the global duplex structure [52]. We therefore assume that 8-oxo-G can be treated as a G in calculations of hybridization free energy.

Inserting these free energies into equation (4.4) with $T = 26$°C yields off-rates, which can be implemented in the LIDA kinetic system (see electronic supplementary material, equations (A 1)–(A 15) in §C). 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 $10 \times 10^{-6}\,\text{M} = 14 \times 10^{-9}\,\text{M} \times \exp(k_{rep} \times \tau_{10\%})$ as $k_{rep} = \ln[(10/14) \times 10^3]/\tau_{10\%}$. The doubling or replication time

[2]A non-catalysed reaction via pseudo blunt-ended ligation may also occur, but much less frequent than the catalysed reaction [44]. Furthermore, in the currently lab-achieved DNA amplification by LIDA, the system is also catalysed by enzymes as noted in the electronic supplementary material of Alladin-Mustan *et al.* [44].

τ_{rep} can be estimated from $A\exp(k_{rep}\times\tau_{rep})=2A$ so $\tau_{rep}=\ln 2/k_{rep}$. Figure 4 right panel shows a histogram of the replication rates obtained for the 160 strands.

The estimated replication rate constants k_{rep} are shown in the right panel of figure 4. They exhibit multiple semi-distinct clusters, with typical doubling times ranging between about 10 min and a couple of hours, with the slowest taking about 3.8 h. Note that this is faster than the estimated vesicle doubling time of 4.3 h, as we found in §3c.

Finally, we need to check the above estimated replication rate constants with the imidazole hydrolysis rate constant $k_{hyd-imp}$. This is because DNA sequences with an estimated overall replication rate constant below $k_{hyd-imp}$ are not viable, given that the replication process will be inhibited by the decay of activated oligomers. The imidazole hydrolysis rate constant is measured to be between 4.78×10^{-6} s^{-1} and 1.04×10^{-5} s^{-1} at room temperature and depending on the ion concentration [61].

Since most of our estimated replication rate constants satisfy $k_{rep}>10^{-4}$ s^{-1}, which is an order of magnitude larger than the largest reported hydrolysis rate constants $k_{hyd-imp}\sim10^{-5}$ s^{-1}, as a first approximation we can assume that a protocellular replication is viable if $k_{rep}>10^{-4}$ s^{-1}

5. Connecting results from charge transport and replication simulations

The overall protocellular fitness is impacted both by the metabolic rate constant k_{CT} and by the co-factor replication rate constant k_{rep}. In fact, as shown in Rocheleau *et al*. [30], there is a nontrivial monotonic relationship between overall fitness (rate of division of protocells) and the two rate constants k_{CT} and k_{rep}. Therefore, the data on the CT and replication rate constants from the simulations performed in §§3b and 4b provide us with information about protocellular fitness. This section explores the combined fitness impact of k_{CT} and k_{rep}.

Co-factor charge transfer and replication rates must be above a certain lower bound in order for the protocellular system to survive and grow. 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, given our analysis in §3b, we require that $k_{CT}>50$ s^{-1} to ensure that the CT rate constant is well above the rate-limiting photo-activation rate constant. From our analysis in §4b, we require that $k_{rep}>10^{-4}$ s^{-1} to ensure that the replication rate constant is well above the associated hydrolysis rates. These two conditions ensure that charge-transport and replication processes can support viable protocells.

Further, recall the discussion in §2, where for simplicity we required that 8-oxo-G is located above the ligation site, which disqualifies a number of strands. In the electronic supplementary material, §D, we enumerate the strands from 1 to 160, and show that strand numbers 65–80 and 113–160 are disqualified for that reason (see table S2 in electronic supplementary material, §D for details). Thus, 95 out of the 160 strands satisfy the oxoG above the ligation site condition. In figure 5 (left panel), we show the replication rate k_{rep} versus the CT rate k_{CT} for all 160 examined strands, where the 'viable area' (purple patch) contains 63 viable strands (indicated by rings). See details in electronic supplementary material, §Db.

It should be emphasized that we combine results from the CT and LIDA simulations without taking into consideration any interactions. 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 the full system dynamics.

(a) Functional characterization of co-factors

To further explore the viability transition discussed above, we can in principle rank the combinatorial co-factors $F(k_{rep},k_{CT},...)$ such that their combined replication and charge transport rates are increasing along an axis to form a fitness ranking of the strands. Again, note that a reliable combined co-factor 'fitness' ranking cannot yet be estimated without a complete protocell-environment simulation. However, due to the theoretical interest of defining a bottom-up fitness ranking, we can still use the results of the CT and LIDA simulations to construct such a co-factor ranking.

How to obtain a combined fitness expression from the charge transfer and the replication process determined by the rate constants k_{CT} and k_{rep}? In Rocheleau *et al*. [30], we studied a similar protocellular system that combined information replication and container growth. This study showed that due to non-trivial couplings between interacting processes, the combined protocellular growth factor is proportional to $[k_{rep}^2 k_{CT}]^{1/3}$. This is shown when the coupled equations (4.1)–(4.6) in Rocheleau *et al*. [30] 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 Rocheleau *et al*. [30], the two systems are closely related and we believe that the results from Rocheleau *et al*. [30] provide a reasonable ansatz for a combined fitness estimate given our current knowledge of these systems. This fitness estimate is shown in the right panel of figure 5, which can be interpreted as a simple approximation for a ranking of protocellular co-factor fitness. Note, however, that the growth law found in Rocheleau *et al*. [30] 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 ranked fitness 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 §3 and figure 3 left panel) as the CT rates vary across more orders of magnitude than the replication rates.

Another suitable way to explore the viability of the presented protocellular system is to estimate the functional information of the involved co-factors [41,42]. This line of investigation is presented and discussed in the electronic supplementary material,

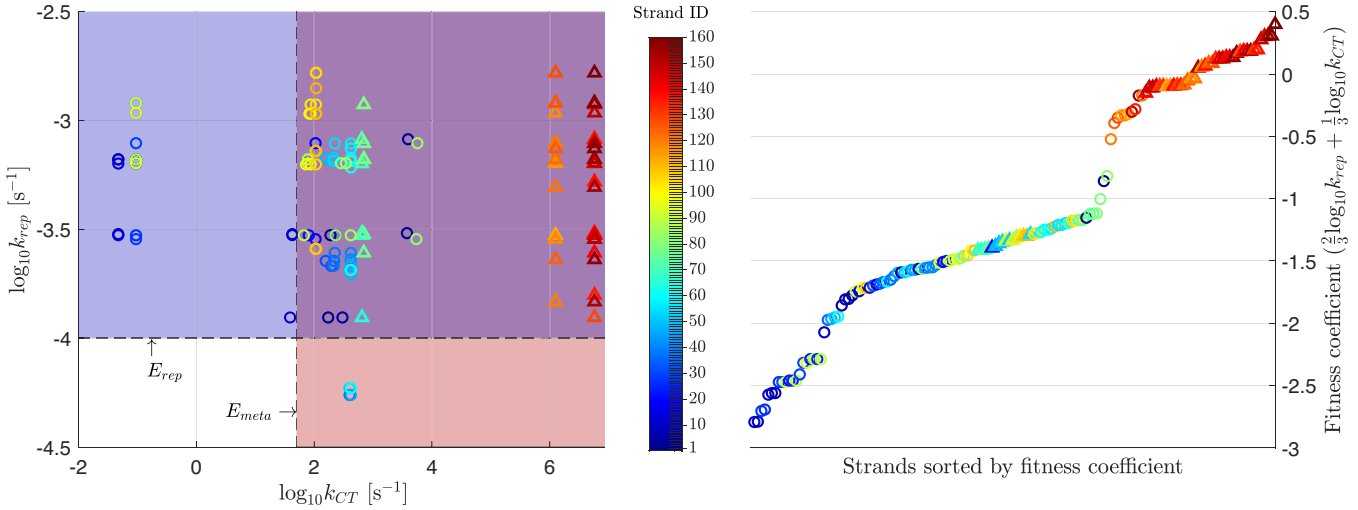


Figure 5. Two ways of combining the previously obtained results for k_{CT} and k_{rep} for all 160 strands. On both figures, each data point corresponds to a unique co-factor, and is coloured to fit its corresponding strand ID (1-160), as shown on the colourbar. The stand ID can be used to obtain the sequence of the strand in the electronic supplementary material, table S2. *Left:* The replication rate k_{rep} versus the CT rate k_{CT} for all 160 examined strands. As mentioned, the colourbar 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 ($k_{CT} > 50$ s^{-1}) and pale blue areas ($k_{rep} > 10^{-4}$ s^{-1}), which were concluded from our investigations in §§3 and 4. Thus, the purple area indicates both viable metabolic and replication rates (both $k_{CT} > 50$ s^{-1} and $k_{rep} > 10^{-4}$ s^{-1}). See text for details. *Right:* Ansatz for a combined metabolic and replication growth rate from Rocheleau *et al.* [30] depicted as the sorted sum of $\frac{2}{3}\log_{10}[k_{rep}] + \frac{1}{3}\log_{10}[k_{CT}]$, as explained in §5*a*. 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 interactions are included in our simulations, so the graph is a simple approximation of the protocellular fitness based on a superposition of k_{CT} and k_{rep} for the 160 examined strands. The colourbar shows which strand corresponds to which data point; rings indicate that oxoG is above the ligation site, while triangles indicate the oxoG is located below the ligation site. See text for details.

§D, where we find that out of the possible 7.30 bits of functional information only about 1.34 bits are required for a protocell to be viable when its co-factor is randomly drawn from the limited available set of 160 co-factors. Intuitively, this is in line with what we observe on figure 5 left panel; many of the tested strands satisfy $k_{CT} > 50$ s^{-1} and $k_{rep} > 10^{-4}$ s^{-1} (63 out of 160).

6. Conclusion

We report results on properties of co-factor molecules that directly regulate the protocellular metabolism. These molecules exhibit two important functions: they serve as part of an electron relay for an energy transduction mechanism, and they act as combinatorial molecules that can replicate and thereby support a primitive form of inheritable information transfer.

The main result of our investigation emerges as we compare the estimated rate constants for charge transport and replication k_{CT} and k_{rep} relative to relevant degradation rates. This comparison indicates that protocellular viability cannot be rejected for 63 of the 160 tested co-factors, given the assumptions that underpin our analysis.

The work is based on an operational definition of minimum life supported by a metabolic energy transducer, an informational co-factors, and a container, which must all be situated in an appropriate environment (see [1,28,37]). 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 capable of replication, and the energy transducer is a ruthenium complex. We present simulation results for CT and replication abilities for 160 DNA co-factors each composed of 18 base pairs [1,28,34–37]. An extensive discussion of the assumptions underpinning the presented simulation investigations and results is found in the electronic supplementary material.

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} through the strands range from 10^{-1} s^{-1} to 10^{6} s^{-1} 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 constant in a system with a direct ruthenium to 8-oxo-G charge transfer (without an intermediate DNA charge transfer) is 1.3×10^{-5} s^{-1}, corresponding to a fatty acid (and vesicle) doubling time of about 4.3 h [34,35,39,59]. If we require the DNA CT to slow down this rate no more than 1%, we require that $k_{CT} > 50$ s^{-1}.

Replication dynamics of the combinatorial DNA co-factor is estimated by LIDA simulations using reaction kinetics ODEs, where reaction constants are determined directly from thermodynamic calculations and observed non-enzymatic ligation rate constant [36]. The obtained overall replication constants k_{rep} typically range from just below 10^{-4} s^{-1} to above 10^{-3} s^{-1}, which means co-factor doubling times from below 20 min to many hours. Oligomer degradation rate constants are observed to be up to about 10^{-5} s^{-1} (imidazole hydrolysis), so we require that $k_{rep} > 10^{-4}$ s^{-1} to ensure viable co-factor replication.

Finally, we estimate co-factor fitness by combining results from the CT and replication (LIDA) simulations. Due to non-trivial interactions between replication and charge transfer processes, the combined protocellular growth factor is estimated to be proportional to $[k_{rep}^2 k_{CT}]^{1/3}$, which defines the co-factor fitness as a function of the replication and charge transport rate constants.

The complete integration of metabolism, information and container into a functional protocell is arguably the greatest challenge to assembling bottom-up minimal living systems. The presented work provides a detailed simulation study of how information can be an integral part of a metabolism, where both the energy transduction and the information complex are kept together on the surface of a container. This work was originally motivated by the question of whether it is experimentally feasible to combine information and metabolism through the proposed energy transducer and co-factor coupling. We conclude that such a coupling should be experimentally feasible, and future laboratory work may test whether our assumptions and predictions are reasonable.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. Supplementary material is available online [62].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. K.R.T.: data curation, formal analysis, investigation, methodology, software, visualization, writing—original draft, writing—review and editing; A.K.: conceptualization, formal analysis, investigation, methodology, validation, writing—review and editing; S.R.: conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualization, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. K.R.T. and S.R. were in part supported by the Donostia International Physics Center (DIPC). A.K. 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.

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.

References

1. Rasmussen S, Bedau MA, Chen L, Deamer D, Krakauer DC, Packard NH, Stadler PF. 2008 *Protocells: bridging nonliving and living matter*. Cambridge, MA: MIT Press. (doi:10.7551/mitpress/9780262182683.001.0001)
2. Adamala KP, Engelhart AE, Szostak JW. 2016 Collaboration between primitive cell membranes and soluble catalysts. *Nat. Commun*. **7**, 11041. (doi:10.1038/ncomms11041)
3. Furubayashi T, Nakano T, Eckford A, Okaie Y, Yomo T. 2016 Packet fragmentation and reassembly in molecular communication. *IEEE Trans. NanoBioscience* **15**, 284–288. (doi:10.1109/tnb.2016.2526051)
4. Kitson PJ, Glatzel S, Chen W, Lin CG, Song YF, Cronin L. 2016 3D printing of versatile reactionware for chemical synthesis. *Nat. Protoc*. **11**, 920–936. (doi:10.1038/nprot.2016.041)
5. Qiao Y, Li M, Booth R, Mann S. 2017 Predatory behaviour in synthetic protocell communities. *Nat. Chem*. **9**, 110–119. (doi:10.1038/nchem.2617)
6. Rodríguez-Arco L, Li M, Mann S. 2017 Phagocytosis-inspired behaviour in synthetic protocell communities of compartmentalized colloidal objects. *Nat. Mater*. **16**, 857–863. (doi:10.1038/nmat4916)
7. Imai M, Sakuma Y, Kurisu M, Walde P. 2022 From vesicles toward protocells and minimal cells. *Soft Matter* **18**, 4823–4849. (doi:10.1039/d1sm01695d)
8. Sato W, Rasmussen M, Deich C, Engelhart AE, Adamala KP. 2022 Expanding luciferase reporter systems for cell-free protein expression. *Sci. Rep.* **12**, 11489. (doi:10.1038/s41598-022-15624-6)
9. Schwille P *et al*. 2018 MaxSynBio: avenues towards creating cells from the bottom up. *Angew. Chem. Int. Ed*. **57**, 13382–13392. (doi:10.1002/anie.201802288)
10. Jahnke K, Weiss M, Frey C, Antona S, Janiesch J, Platzman I, Göpfrich K, Spatz JP. 2019 Programmable functionalization of surfactant-stabilized microfluidic droplets via DNA-tags. *Adv. Funct. Mater*. **29**, 1808647. (doi:10.1002/adfm.201808647)
11. Zwart H. 2019 From primal scenes to synthetic cells. *eLife* **8**, e46518. (doi:10.7554/elife.46518)
12. Pols T, Sikkema HR, Gaastra BF, Frallicciardi J, Śmigiel WM, Singh S, Poolman B. 2019 A synthetic metabolic network for physicochemical homeostasis. *Nat. Commun*. **10**, 4239. (doi:10.1038/s41467-019-12287-2)
13. Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA III, Smith HO, Venter JC. 2006 Essential genes of a minimal bacterium. *Proc. Natl Acad. Sci. USA* **103**, 425–430. (doi:10.1073/pnas.0510013103)
14. Villani M, Filisetti A, Graudenzi A, Damiani C, Carletti T, Serra R. 2014 Growth and division in a dynamic protocell model. *Life* **4**, 837–864. (doi:10.3390/life4040837)
15. Shirt-Ediss B, Solé R, Ruiz-Mirazo K. 2015 Emergent chemical behavior in variable-volume protocells. *Life* **5**, 181–211. (doi:10.3390/life5010181)
16. Farmer JD, Kauffman SA, Packard NH. 1986 Autocatalytic replication of polymers. *Physica D* **22**, 50–67.
17. Vasas V, Fernando C, Santos M, Kauffman S, Szathmáry E. 2012 Evolution before genes. *Biol. Direct* **7**, 1–14. (doi:10.1186/1745-6150-7-1)
18. Tanaka S, Fellermann H, Rasmussen S. 2014 Structure and selection in an autocatalytic binary polymer model. *Europhys. Lett.* **107**, 28004. (doi:10.1209/0295-5075/107/28004)
19. Adamala K, Szostak JW. 2013 Nonenzymatic template-directed RNA synthesis inside model protocells. *Science* **342**, 1098–1100. (doi:10.1126/science.1241888)
20. Sievers D, von Kiedrowski G. 1998 Self-replication of hexadeoxynucleotide analogues: autocatalysis versus cross-catalysis. *Chem. Eur. J*. **4**, 629–641. (doi:10.1002/(sici)1521-3765(19980416)4:43.0.co;2-0)
21. Lincoln TA, Joyce GF. 2009 Self-sustained replication of an RNA enzyme. *Science* **323**, 1229–1232. (doi:10.1126/science.1167856)
22. Segré D, Ben-Eli D, Lancet D. 2000 Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. *Proc. Natl Acad. Sci. USA* **97**, 4112–4117. (doi:10.1073/pnas.97.8.4112)
23. Fellermann H, Corominas-Murtra B, Hansen PL, Ipsen JH, Solé R, Rasmussen S. 2015 Non-equilibrium thermodynamics of self-replicating protocells. *arXiv* 1503.04683. (doi:10.48550/arXiv.1503.04683)
24. Fellermann H, Markovitch O, Gilfellon O, Madsen C, Phillips A. 2016 Toward programmable biology. *ACS Synth. Biol*. **5**, 793–794. (doi:10.1021/acssynbio.6b00213)
25. Corominas-Murtra B. 2019 Thermodynamics of duplication thresholds in synthetic protocell systems. *Life* **9**, 9. (doi:10.3390/life9010009)
26. England JL. 2013 Statistical physics of self-replication. *J. Chem. Phys.* **139**, 09B623. (doi:10.1063/1.4818538)

27. Kolchinsky A. 2024 Thermodynamic dissipation does not bound replicator growth and decay rates. *arXiv* 2404.01130. (doi:10.48550/arXiv.2404.01130)
28. Rasmussen S, Chen L, Nilsson M, Abe S. 2003 Bridging nonliving and living matter. *Artif. Life* **9**, 269–316. (doi:10.1162/106454603322392479)
29. Carletti T, Serra R, Poli I, Villani M, Filisetti A. 2008 Sufficient conditions for emergent synchronization in protocell models. *J. Theor. Biol*. **254**, 741–751. (doi:10.1016/j.jtbi.2008.07.008)
30. Rocheleau T, Rasmussen S, Nielsen PE, Jacobi MN, Ziock H. 2007 Emergence of protocellular growth laws. *Phil. Trans. R. Soc. B* **362**, 1841–1845. (doi:10.1098/rstb.2007.2076)
31. Rasmussen S, Bedau MA, John S, McCaskill NHP. 2008 A roadmap to protocells. In *Protocells: bridging nonliving and living matter* (eds S Rasmussen, MA Bedau, L Chen, D Deamer, DC Krakauer, NH Packard, PF Stadler), pp. 71–100. Cambridge, MA: MIT Press. (doi:10.7551/mitpress/7590.003.0008)
32. Rasmussen S, Chen L, Deamer D, Krakauer DC, Packard NH, Stadler PF, Bedau MA. 2004 Transitions from nonliving to living matter. *Science* **303**, 963–965. (doi:10.1126/science.1093669)
33. Fellermann H, Rasmussen S, Ziock HJ, Solé RV. 2007 Life cycle of a minimal protocell—a dissipative particle dynamics study. *Artif. Life* **13**, 319–345. (doi:10.1162/artl.2007.13.4.319)
34. DeClue MS, Monnard PA, Bailey JA, Maurer SE, Collis GE, Ziock HJ, Rasmussen S, Boncella JM. 2009 Nucleobase mediated, photocatalytic vesicle formation from an ester precursor. *J. Am. Chem. Soc*. **131**, 931–933. (doi:10.1021/ja808200n)
35. Maurer SE, DeClue MS, Albertsen AN, Dörr M, Kuiper DS, Ziock H, Rasmussen S, Boncella JM, Monnard PA. 2011 Interactions between catalysts and amphiphilic structures and their implications for a protocell model. *ChemPhysChem* **12**, 828–835. (doi:10.1002/cphc.201000843)
36. Cape JL, Edson JB, Spencer LP, DeClue MS, Ziock HJ, Maurer S, Rasmussen S, Monnard PA, Boncella JM. 2012 Phototriggered DNA phosphoramidate ligation in a tandem 5⊠-amine deprotection/3⊠-imidazole activated phosphate coupling reaction. *Bioconjugate Chem*. **23**, 2014–2019. (doi:10.1021/bc300093y)
37. Rasmussen S, Constantinescu A, Svaneborg C. 2016 Generating minimal living systems from non-living materials and increasing their evolutionary abilities. *Phil. Trans. R. Soc. B* **371**, 20150440. (doi:10.1098/rstb.2015.0440)
38. Engelhardt J, Andresen B, Rasmussen S. 2020 Thermodynamics and kinetics of Lesion Induced DNA Amplification (LIDA). In *ALIFE 2020: The 2020 Conf. on artificial life*, pp. 269–272. Cambridge, MA: MIT Press. (doi:10.1162/isal_a_00312)
39. Bornebusch DH, Ziock HJ, Rasmussen S. 2021 Reaction kinetics exploration of a protocellular metabolism (eds J Čejková, S Holler, L Soros, O Witkowski). In *ALIFE 2021: The 2021 Conf. on artificial life*. Cambridge, MA: MIT Press. (doi:10.1162/isal_a_00389)
40. Thomsen KR. 2022 *Energy, information and the emergence of living matter*. Odense, Denmark: University of Southern Denmark.
41. Szostak JW. 2003 Functional information: molecular messages. *Nature* **423**, 689–689. (doi:10.1038/423689a)
42. Hazen RM, Griffin PL, Carothers JM, Szostak JW. 2007 Functional information and the emergence of biocomplexity. *Proc. Natl Acad. Sci. USA* **104**, 8574–8581. (doi:10.1073/pnas.0701744104)
43. von Kiedrowski G. 1993 Minimal replicator theory i: parabolic versus exponential growth. In *Bioorganic chemistry frontiers* (eds H Dugas, FP Schmidtchen), pp. 113–146, vol. 3. Berlin, Germany: Springer. (doi:10.1007/978-3-642-78110-0_4)
44. Alladin-Mustan BS, Mitran CJ, Gibbs-Davis JM. 2015 Achieving room temperature DNA amplification by dialling in destabilization. *Chem. Commun*. **51**, 9101–9104. (doi:10.1039/c5cc01548k)
45. Voityuk AA. 2005 Estimates of electronic coupling for excess electron transfer in DNA. *J. Chem. Phys*. **123**, 34903. (doi:10.1063/1.1961400)
46. Loft S, Poulsen HE. 1996 Cancer risk and oxidative DNA damage in man. *J. Mol. Med*. **74**, 297–312. (doi:10.1007/s001090050031)
47. Roszkowski K, Jozwicki W, Blaszczyk P, Mucha-Malecka A, Siomek A. 2011 Oxidative damage DNA: 8-oxoGua and 8-oxodG as molecular markers of cancer. *Med. Sci. Monit*. **17**, CR329–CR333. (doi:10.12659/msm.881805)
48. Drummond TG, Hill MG, Barton JK. 2003 Electrochemical DNA sensors. *Nat. Biotechnol.* **21**, 1192–1199. (doi:10.1038/nbt873)
49. Mihailovic A, Vladescu I, McCauley M, Ly E, Williams MC, Spain EM, Nuñez ME. 2006 Exploring the interaction of ruthenium(II) polypyridyl complexes with DNA using single-molecule techniques. *Langmuir* **22**, 4699–4709. (doi:10.1021/la053242r)
50. Genereux JC, Barton JK. 2010 Mechanisms for DNA charge transport. *Chem. Rev*. **110**, 1642–1662. (doi:10.1021/cr900228f)
51. Williams TT, Odom DT, Barton JK. 2000 Variations in DNA charge transport with nucleotide composition and sequence. *J. Am. Chem. Soc*. **122**, 9048–9049. (doi:10.1021/ja001552k)
52. Gasper SM, Schuster GB. 1997 Intramolecular photoinduced electron transfer to anthraquinones linked to duplex DNA: the effect of gaps and traps on long-range radical cation migration. *J. Am. Chem. Soc*. **119**, 12762–12771. (doi:10.1021/ja972496z)
53. Bixon M, Jortner J. 2002 Long-range and very long-range charge transport in DNA. *Chem. Phys*. **281**, 393–408. (doi:10.1016/s0301-0104(02)00495-0)
54. Giese B, Amaudrut J, Köhler AK, Spormann M, Wessely S. 2001 Direct observation of hole transfer through DNA by hopping between adenine bases and by tunnelling. *Nature* **412**, 318–320. (doi:10.1038/35085542)
55. Osakada Y, Kawai K, Fujitsuka M, Majima T. 2008 Kinetics of charge transfer in DNA containing a mismatch. *Nucleic Acids Res*. **36**, 5562–5570. (doi:10.1093/nar/gkn505)
56. SantaLucia J, Hicks D. 2004 The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct*. **33**, 415–440. (doi:10.1146/annurev.biophys.32.110601.141800)
57. Bhattacharya PK, Barton JK. 2001 Influence of intervening mismatches on long-range guanine oxidation in DNA duplexes. *J. Am. Chem. Soc*. **123**, 8649–8656. (doi:10.1021/ja010996t)
58. Bornebusch DH. 2021 Reaction kinetic exploration of a protocellular metabolism and information. Master's thesis, University of Southern Denmark, Odense.
59. Polizzi NF, Therien MJ, Beratan DN. 2016 Mean first-passage times in biology. *Isr. J. Chem*. **56**, 816–824. (doi:10.1002/ijch.201600040)
60. Christensen U, Jacobsen N, Rajwanshi VK, Wengel J, Koch T. 2001 Stopped-flow kinetics of locked nucleic acid (LNA)oligonucleotide duplex formation: studies of LNA-DNA and DNA-DNA interactions. *Biochem. J.* **354**, 481–484. (doi:10.1042/0264-6021:3540481)
61. Kawamura K, Maeda J. 2007 Kinetic analysis of oligo(C) formation from the 5′-monophosphorimidazolide of cytidine with Pb(II) ion catalyst at 10-75 degrees C. *Orig. Life Evol. Biosph.* **37**, 153–165. (doi:10.1007/s11084-006-9063-0)
62. Reinholt Thomsen K, Kolchinsky A, Rasmussen S. 2025 Supplementary material from: Protocellular energy transduction, information, and fitness. Figshare. (doi:10.6084/m9.figshare.c.8051168)