SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
BIOPHYSICAL CHEMISTRY
COMPUTATIONAL CHEMISTRY

Received
15 September 2014

Accepted
22 December 2014

Published
26 January 2015

Correspondence and requests for materials should be addressed to D.Y.Z. (zubarev@fas.harvard.edu) or A.A.-G. (aspuru@chemistry.harvard.edu)

Uncertainty of Prebiotic Scenarios: The Case of the Non-Enzymatic Reverse Tricarboxylic Acid Cycle

Dmitry Yu Zubarev, Dmitrij Rappoport & Alán Aspuru-Guzik

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA.

We consider the hypothesis of the primordial nature of the non-enzymatic reverse tricarboxylic acid (rTCA) cycle and describe a modeling approach to quantify the uncertainty of this hypothesis due to the combinatorial aspect of the constituent chemical transformations. Our results suggest that a) rTCA cycle belongs to a degenerate optimum of auto-catalytic cycles, and b) the set of targets for investigations of the origin of the common metabolic core should be significantly extended.

The current diversity of hypotheses related to the chemical origin of life[1] is a result of the fundamental lack of the detailed information about the physical-chemical conditions on early Earth. Without such information it is problematic to evaluate the viability of reaction channels associated with any hypothetical reaction sequence and discriminate between hypotheses. Reduction of the uncertainty in the global and local conditions that facilitated prebiotic reactions remains a formidable challenge. In this paper, we take a different route and attempt to define and evaluate uncertainty of hypothetical prebiotic scenarios that cannot be reduced until external conditions are precisely characterized. Specifically, the hypothesis about the predisposed nature of the non-enzymatic reverse tricarboxylic acid (rTCA) cycle under prebiotic conditions is considered[2,3]. In the following sections we show how the hypothesis can be recast into a chemical reaction network model and how this model can be evaluated using a combination of symbolic and quantum chemistry. This evaluation yields a network representation of the set of potential transformations that we use to gain insights into the issue of the diversity of the chemical routes that are consistent with the hypothetical chemistry of the non-enzymatic rTCA cycle.

The structure of the enzymatic form of the rTCA cycle, as found in several groups of anaerobic bacteria[4,5], is shown in Figure 1 along with the 7 reaction types, to which the reactions of the cycle belong. The non-enzymatic version of the rTCA cycle has been hypothesized to be central to the origin of life and have unique properties[2]. The hypothesis was inspired by the results of the data mining in Beilstein database of organic molecules. A set of *ad hoc* selection rules was formulated to prune 3.5 million entries and "to generate the emergence of the reductive citric acid cycle from the master list of compounds"[2]. The rules included a) bounds on the number of C, H, and O atoms in the molecules, b) high water solubility, c) low heats of combustion, d) presence of carbonyl groups, and e) absence of cycles and synthetically inaccessible groups, such as ethers, triple CC bonds, and peroxy groups. It was found that the 11 molecules of the rTCA cycle are within the group of 153 molecules in the pruned dataset. This was taken as an "indication that the chemistry at the core of the metabolic chart is necessary and deterministic"[2]. The follow-up study[3] further developed the hypothesis of the universal character and primordial relevance of the non-enzymatic rTCA cycle by evaluating its concentration dependence and network topology. It was argued, that a) the autocatalytic structure over a single short loop contributes to the cycle self-enhancement at the low level of complexity, and b) redundancy of the synthetic steps reduces the number of innovations required for cycle discovery in the pre-enzymatic world[3]. Also, the argument about the favorable relationships between free energies of formation and redox potential of the rTCA cycle molecules was made to support the idea of their inevitable emergence on the arbitrary prebiotic relaxation pathways.

The hypothesis was criticized on multiple grounds. It was pointed out that empirical chemical databases are strongly biased towards natural product chemistry, whose central component is the forward TCA cycle[6,7]. Therefore, the data mining attempts were bound to find the molecules of the rTCA cycle. The following problems with the data mining strategy were identified: a) the pruning rules can be too restrictive, b) the organic chemists could have little interest in the molecules allowed by the rules but without significant biochemical significance,

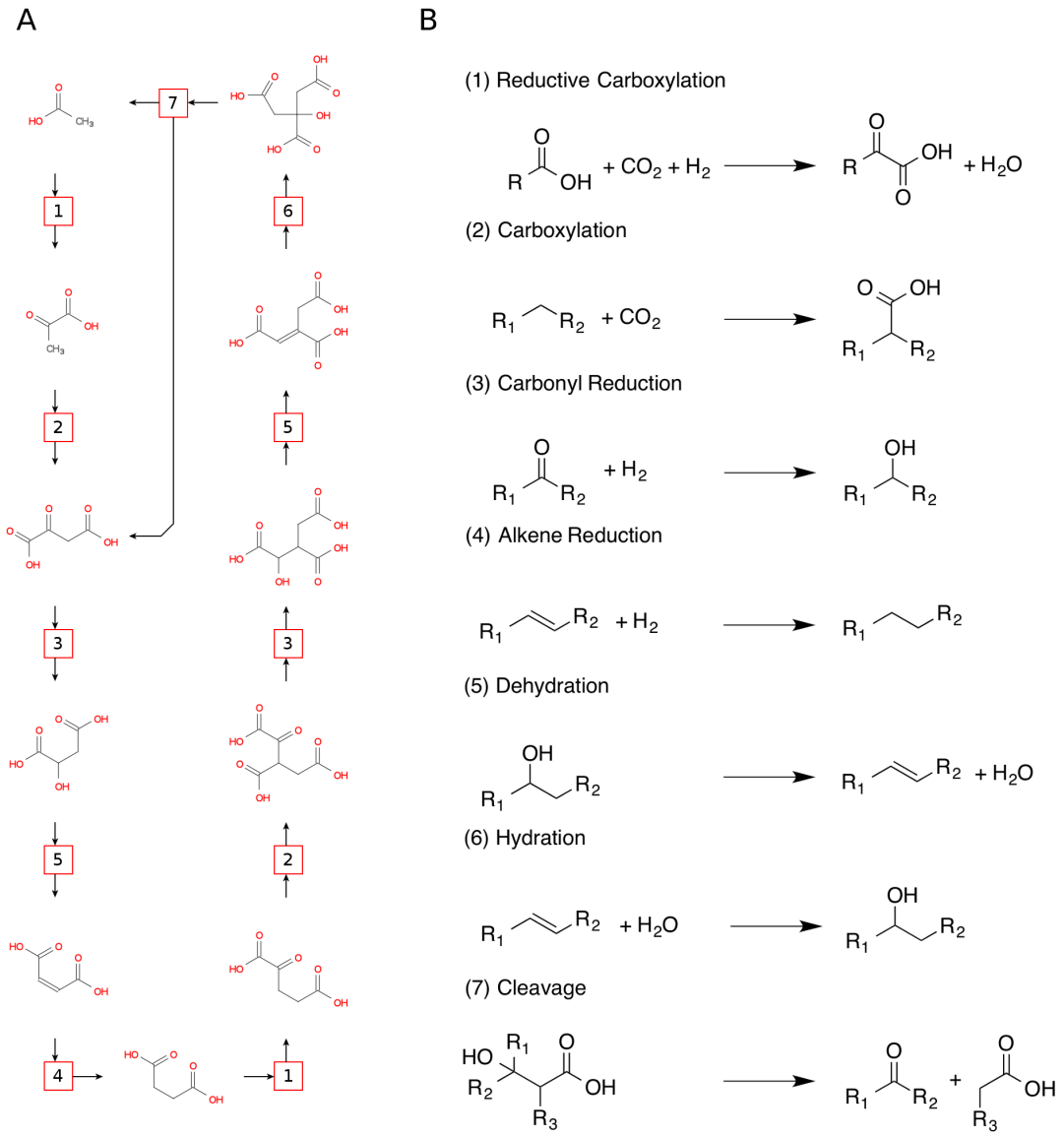

Figure 1 | **Structure of the rTCA cycle.** Panel A: sequence of the substrates and reactions. Reactions are labeled according to the reaction types described on panel B. The autocatalytic structure of the cycle derives from the branching point associated with citrate cleavage.

and c) it can be difficult for synthetic chemistry to produce many non-biological compounds. The first two issues make it problematic to consider the hypothesis prebiotically relevant[6]. The thermodynamic viability of the transformations was also questioned[8]. Nevertheless, experimental efforts were made to investigate chemistries that might have produced the components of the non-enzymatic rTCA cycle[9–11].

The rTCA cycle that is found in bacteria is catalyzed by enzymes with high degrees of substrate selectivity. While different phylogenetic groups show some variations in their enzymatic machinery, the reaction substrates and the reaction sequence of the enzymatic rTCA cycle are conserved[5]. On the other hand, the transformations of prebiological chemistry are assumed to occur under the effect of chemical catalysts. The latter, however, are typically active with respect to certain *types of chemical transformations* and lack the high substrate selectivity characteristic of enzyme catalysts. Variations of the external conditions, e.g., heating/cooling cycles, pH, or exposure to UV radiation, can trigger different reaction sequences consistent with the respective reaction types (Fig. 1). The union of these reaction sequences can be described as a chemical supernetwork, which is solely defined by the admissible *reaction types*. Given detailed mechanistic information along with the external conditions, it should be possible to evaluate the reaction rates, yields, time-dependent concentration profiles up to the equilibration, and reduce the chemical supernetwork to one or few significant reaction networks. Therefore, the supernetwork structure expresses the uncertainty in the reaction parameters and external conditions inherent in the model of chemistry. The range of the structural and energetic parameters on the supernetwork in turn quantifies the uncertainty of the proposed model.

In this study, we construct a chemical supernetwork consistent with the reaction types of the rTCA cycle. A schematic representation of the process is shown in Fig 2A and explained in details in Methods section. The reaction types of the reactivity model (Fig. 1B) are encoded as symbolic transformations. The transformations are iteratively applied to the set of the available molecules that carry appropriate functional groups. The symbolic representations of the reactants are converted into symbolic representations of the products. These symbolically encoded molecular structures are transformed into a representation suitable for quantum chemical calculations. Quantum chemistry is used to validate viability of the molecules, and provide energetic characteristics of the reaction that

has been carried out symbolically (see Methods). If all the molecules in the generated reaction pass the validation, the products are added to the list of the available molecules.

The rTCA cycle is effectively a mechanism of carbon fixation that can be started from any point along the cycle. We chose acetate as the initial substrate because it is the smallest and simplest organic molecule that appears in the rTCA cycle. There is a set of simple inorganic molecules that are available for any reaction at any stage of the supernetwork construction. It includes water, carbon dioxide, and hydrogen, which serves as a proxy for the actual reducing agent. There are 11 steps in the reaction sequence of the rTCA cycle. Therefore, we carried out 11 iterations of the supernetwork construction process and produced the smallest supernetwork that is guaranteed to have the rTCA cycle in it.

Results

Structure of rTCA supernetwork. The smallest supernetwork that includes rTCA cycle is designated the rTCA supernetwork (Fig. 2B). It contains 175 molecules and 444 reactions. Only the organic substrates are considered, the simple molecules H_2, H_2O, and CO_2 are excluded from the dataset. Analysis of the connectivity of molecular nodes (see SI) reveals that the chemistry of glyoxylate is at least as important as the chemistry of acetate. Acetate, which served as a starting point of the reconstruction process, has only the second largest degree with 14 incoming and 3 outgoing connections. Glyoxylate has 18 incoming and 4 outgoing connections. Both molecules are the most frequent products in the network due to cleavage reactions (Fig. 1B). This is a suggestive outcome in view of the existing glyoxylate scenarios[7,12–14]. It indicates that the rTCA cycle hypothesis and glyoxylate scenarios can be merged into a single scenario.

Having constructed the rTCA supernetwork, we are now in the position to evaluate the hypothesis of the uniqueness and predisposed nature of the rTCA cycle within the supernetwork. Specifically, we will enumerate cycles that involve acetate and glyoxylate and compare their properties. Furthermore, one can investigate its evolution following some model, for example, mass-action kinetics, and see what subset of reactions is selected. One would have to obtain rate constants for the supernetwork reactions, specify the initial concentrations, environmental conditions, such as pressure, temperature and acidity, and make assumptions about the availability of catalysts and their activity and selectivity. Such investigations are beyond the scope of the present study.

A total of 1881 cycles containing acetate, glyoxylate, or both is found in the rTCA supernetwork (see SI for details). These are not thermodynamic cycles with a change of a free Gibbs energy identically equal to zero. Rather, they are reaction sequences that recreate one or several substrates that were consumed, cf. an autocatalytic cycle, and have a non-trivial net stoichiometry. For example, the net reaction equation of the rTCA cycle is $4CO_2 + 5H_2 \rightarrow$ oxaloacetate $+ 3H_2O$. Short cycles between two molecules due to hydration/dehydration and carboxylation/decarboxylation, i.e., cycles with less than 6 members, are disregarded. A single branching point forming an autocatalytic loop is present in 758 cycles. There are 174 cycles with 2 branching points and 20 with 3 branching points. The remaining cycles do not contain branching points. Figure 3 shows distributions of the cycles with respect to the following characteristics: i) *length* – the number of molecules in the cycle relative to the number of molecules in rTCA cycle; ii) *overlap* – the ratio between the number of molecules shared with rTCA cycle and the length of rTCA cycle; iii) *cost* - the net change of Gibbs free energy along the cycle; and iv) *height* – the difference between the highest and the lowest reaction Gibbs free energies along the cycle. Gibbs free energies used in (iii) and (iv) are computed in aqueous solution (see Methods).

The shapes of the histograms in Figure 3 are not smooth enough to identify the underlying distributions, but they exhibit distinguishable peaks and tails. The molecular sequence of the rTCA cycle is clearly unique judging by the overlap distribution. The rTCA cycle does not belong to the highly populated regions and sits on the tails of the length and cost distributions, and between two modes of the height distribution. Fig. 3B gives a better perspective on the structure of the distributions of the energetic characteristics of the cycle. It shows

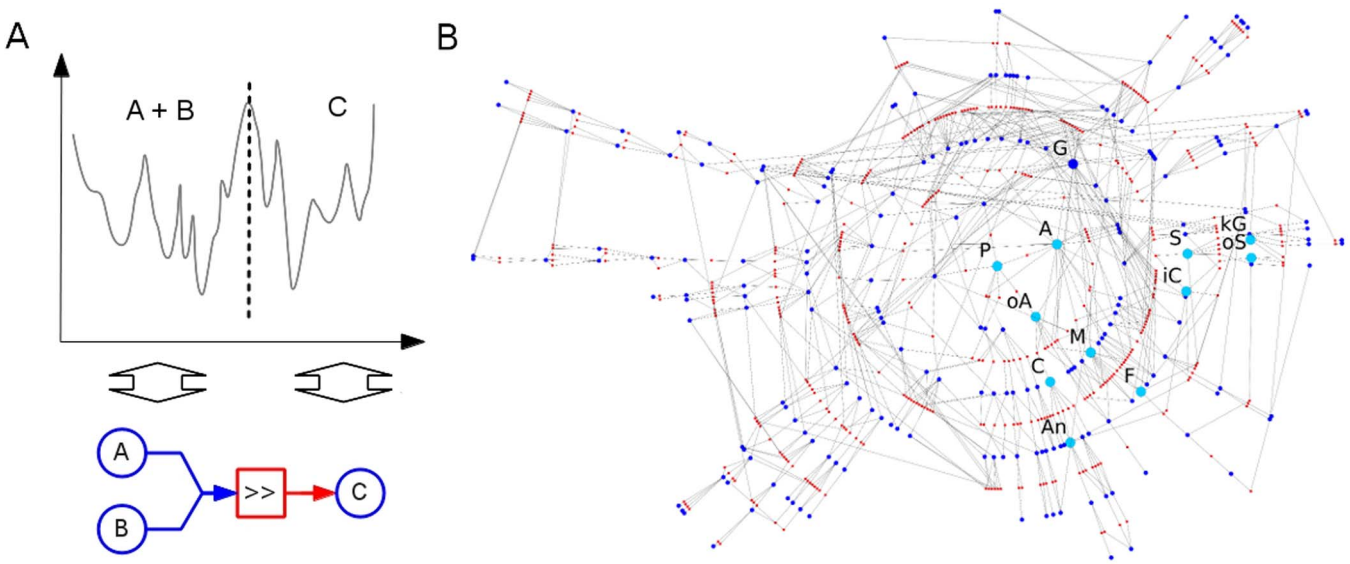

Figure 2 | Construction of the rTCA chemical supernetwork. Panel A: general workflow. Initial molecules "A" and "B" are validated via geometry optimization on a quantum chemical potential energy surface (PES). Their quantum chemical representations are converted into string representations of the respective Lewis structures. The latter are transformed symbolically into reaction product "C" according to a reaction rule "≫". Reaction rules specify mappings between functional groups according to the types of the chemical reactions, i.e., model parameters. Lewis structure "C" is mapped back onto PES and validated via quantum chemical calculations. Panel B: rTCA supernetwork constructed according to the reaction types of Fig. 1B. Red nodes are reactions, smaller (blue) and larger (cyan) nodes are molecules. Larger (cyan) molecular nodes label molecules of rTCA cycle: A – acetate, P – pyruvate, oA – oxaloacetate, M – malate, F – fumarate, S – succinate, kG – ketoglutarate, oS – oxalosuccinate, iC -isocitrate, An – aconitate, C – citrate; G is glyoxylate.

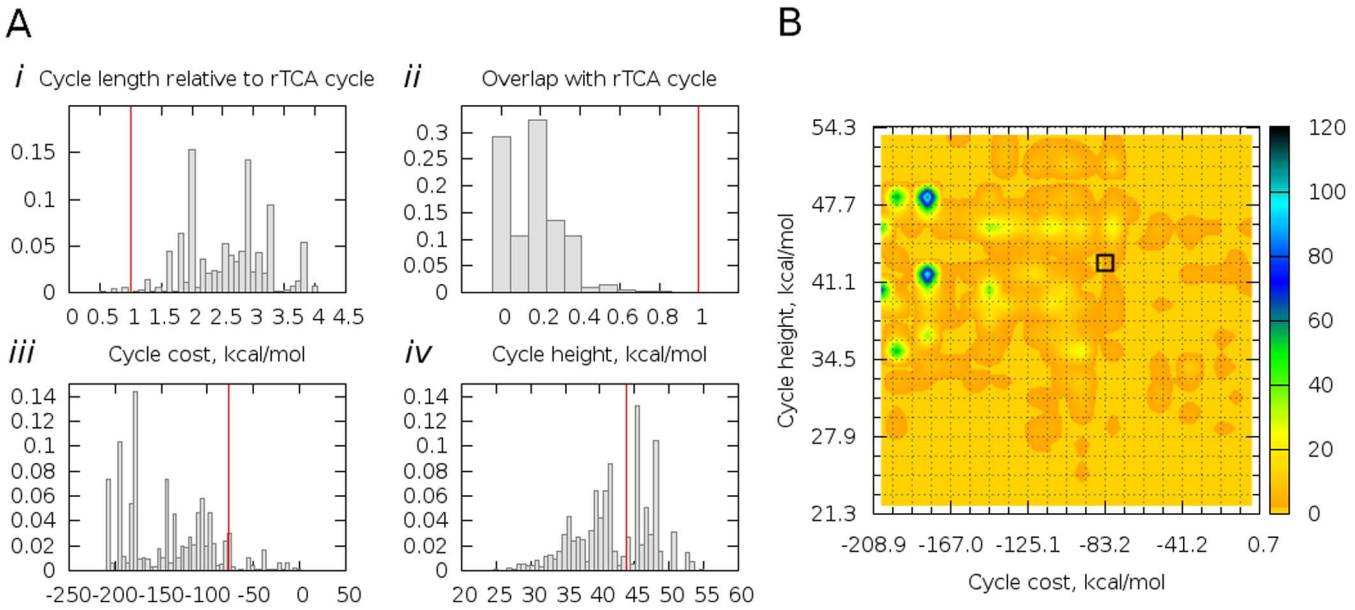

Figure 3 | Statistics of the reductive cycles found in rTCA supernetwork. Thermodynamic values are computed in aqueous solution (see Methods). The vertical red lines and the black square represent the computed values for the rTCA cycle. Panel A: i. Relative cycle *length* defined as a ratio between the number of molecules in the cycle and rTCA cycle; ii. *Overlap* with rTCA cycle defined as a ratio between the number of molecules shared with rTCA cycle and the length of rTCA cycle; iii. Cycle *cost* defined as the net change of Gibbs free energy along the cycle; and iv. Cycle *height* defined as the difference between the highest and the lowest reaction Gibbs free energies along the cycle. Red vertical lines mark position of rTCA cycle. Panel B: heat map of the distribution with respect to the cycle cost and cycle height. Color codes bin occupation.

multiple highly populated domains and the rTCA cycle is found on the outskirts of one of low-populated domains.This suggests that the rTCA cycle has a low probability of realization in a random process, that selects for the respective properties and has access to the entire range of the property values. The cycle characteristics are shifted toward extreme values in length and cost distributions, yet there are enough cycles to compete with rTCA cycle in terms of the extremality of these characteristics.

Redox properties of molecules in the supernetwork. The similar picture emerges from the analysis of the redox properties of the molecules in the supernetwork. Previously-proposed conclusions about the predisposed nature of rTCA cycle[3] in part relied on the analysis of intensive characteristics of the rTCA cycle molecules such as thermodynamic stability per carbon atom and degree of reduction per carbon atom. Following Ref. 3, we define the reducing potential per carbon atom taken from the environment to form molecule A as the ratio y/x determined from a formal reaction $xCO_2 + yH_2 \rightarrow A + zH_2O$. As a counterpart of Figure 2 in Ref. 3, Figure 4A shows theoretical estimates of the free energy of formation computed in aqueous solution (see Methods) per carbon atom plotted against reducing potential for all molecules of the generated supernetwork.

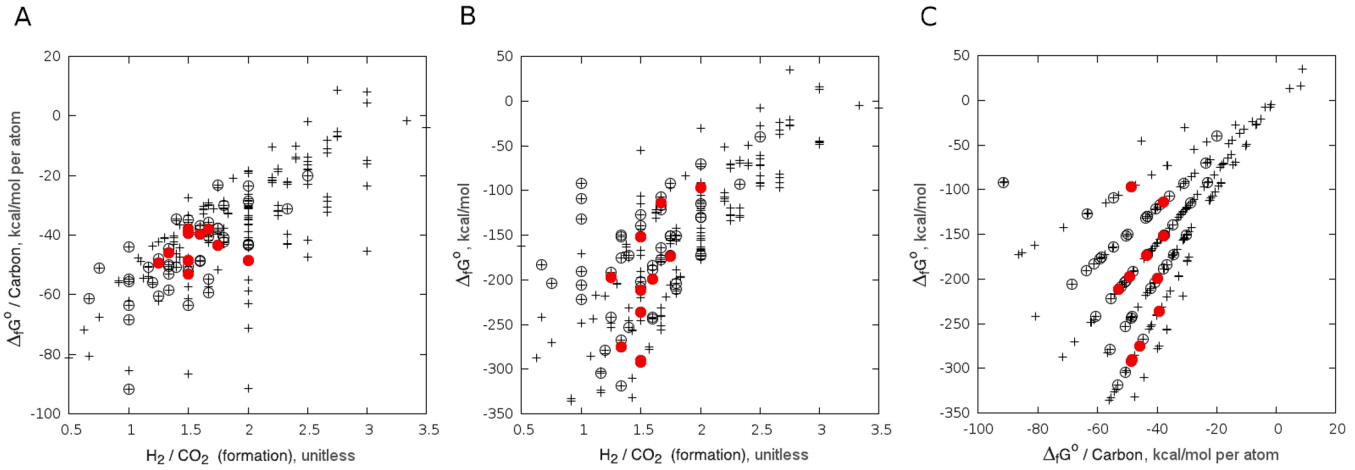

Figure 4 | Stability and reducing potential of the molecules in the rTCA supernetwork. Thermodynamic values are computed in aqueous solution (see Methods). H_2/CO_2 designates reducing potential of molecule A defined as y/x following equation $xCO_2 + yH_2 \rightarrow A + zH_2O$. The crosses represent all molecules in the supernetwork. Empty circles over crosses represent the molecules from the cycles similar to rTCA cycle with 11 molecules and 1 branching point. The filled circles represent the molecules of rTCA cycle. Panel A: A counterpart of the Figure 2 from Ref. 3; computed free energies of formation are used instead of experimental. Free energy of formation per carbon atom is plotted against degree of reduction. Panel B: Free energy of formation vs. degree of reduction. Panel C: Free energy of formation vs. free energy of formation per carbon atom. The dataset is strongly stratified according to the number of carbon atoms from 1 to 7, from the top of the plot.

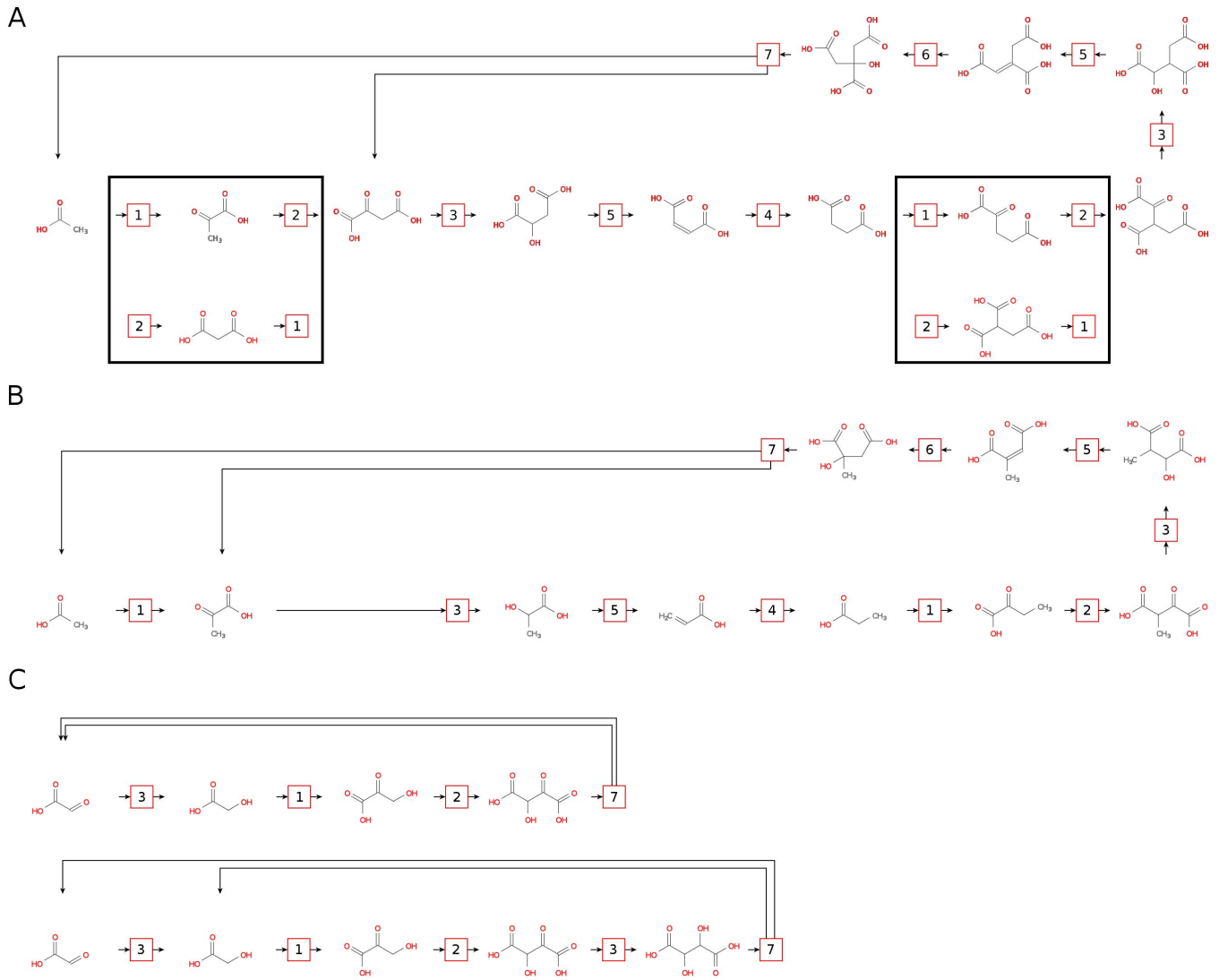

Figure 5 | Examples of auto-catalytic cycles found in the rTCA supernetwork. Panel A: cycles with the fewest modifications in the molecular sequence with respect to the rTCA cycle. Reordering of the reductive carboxylation and carboxylation steps "1" and "2" in two fragments of rTCA cycle (black boxes) yields three auto-catalytic cycles that have the closest molecular sequences to the rTCA. Panel B: a cycle with the fewest modifications in the reaction types sequence with respect to the rTCA cycle. The sequence of reaction types in rTCA cycle is "1 2 3 5 4 1 2 3 5 6 7" starting from the reductive carboxylation of acetate (See Fig. 1). Omission of the first carboxylation step "2" involving pyruvate leads to the cycle with the sequence "1 _ 3 5 4 1 2 3 5 6 7" with a very different molecular composition. Panel C: two cycles involving the molecules with the highest closeness centrality in rTCA network. Both sequences are based on glyoxylate.

Molecules of rTCA cycle belong to the central part of a broad distribution and have neighbors with comparable or more favorable combinations of stability and reducing potential. For example, the cycles obtained by permutations of reductive carboxylation and carboxylation steps (Fig. 5A) would qualify as more favorable on thermodynamic grounds. They proceed via malonate which is more stable than pyruvate with −176 vs. −114 kcal/mol per 3 carbon atoms, and/or 1, 1, 2-ethanetricarboxylate that is more stable than 2-ketoglutarate with −254 vs. −199 kcal/mol per 5 carbon atoms.

Intensive characteristics are convenient for comparison of systems of different sizes, such as substrates with different numbers of carbon atoms. Evaluation of viability of chemical processes typically relies on an extensive property, the free energy of formation in the current context. We reveal a more detailed structure of the distribution of Fig. 4A by adding this property as the third dimension and considering two projections as shown in Figs. 4B and 4C. The position of rTCA cycle in Fig. 4A is within 1.25 to 2.0 dimensionless units interval along reducing potential axis and in the −52.9 to −37.9 kcal/mol interval along free energy of formation per carbon atom axis. Figs. 4B and 4C show that these intervals indeed contain some of the most stable molecules, but the most stable molecules do not belong to rTCA cycle. Figure 4C clearly illustrates this statement because of the strong stratification of the dataset. Each layer contains molecules with a fixed number of carbon atoms from 1 to 7. Molecules of rTCA cycle are found close to the high-stability ends of the layers and have several more stable neighbors encountered in alternative cycles with structural characteristics comparable to the rTCA cycle.

Discussion

The main objective of this work is to offer a measure of uncertainty quantification for the rTCA prebiotic scenario. We use reactions of the rTCA cycle to define a model of reactivity in operational terms. This uncertainty is less straightforward to define than in traditional approaches of uncertainty quantification[15]. In order to assess the uncertainty, we use a set of well-defined and chemically relevant

network characteristics. Thus we are able to quantify the dispersion of the characteristics and approach the question whether the rTCA cycle is predisposed by its chemistry.

Considering the shapes of the distributions constructed in the previous section, we conclude that the rTCA cycle should have a low probability of a random realization. We also notice that its length and cost are close to the extreme values. Selection for the extreme values implies an optimization process. Is there any evidence so far that such optimization will inevitably lead to the rTCA cycle? The energy-based distributions suggest that there are multiple cycles with characteristics very similar to the rTCA cycle. The natural variance in the external conditions implies imperfect selectivity for any conceivable non-biological optimization process. Such a process could reduce the number of the available molecules and cycles built on them, but it would not be able to yield rTCA cycle exclusively. Therefore, we conclude that the rTCA cycle belongs to a degenerate optimum – a set of multiple cycles with optimal, i.e., extreme, characteristics. Further selection into biological cycles may have occurred by other means, such as a frozen accident, that is, the selection and preservation of a particular pathway from the ensemble of possibilities due to an undetermined random event[16]. Uniqueness of the molecular sequence of the rTCA cycle (overlap distribution, Fig. 3A) hints evolution of the catalytic selectivity as the factor responsible for this particular selection.

Evaluation of the uncertainty of the hypothesis regarding the origin of the common metabolic core[2,3] inevitably changes perception of the targets for experimental studies. In addition to rTCA cycle itself, one can investigate the viability of the related cycles. For example, reordering of the reductive carboxylation and carboxylation steps in two sections of the rTCA cycle as shown in Figure 5A yields its closest relatives in terms of the molecular sequences. The new cycles proceed via malonate instead of pyruvate, or 1, 1, 2-ethanetricarboxylate instead of 2-ketoglutarate, or both. Further, we note that non-enzymatic nature of the described hypothetical chemistry implies low selectivity of the available catalysts with respect to the substrates. If the specific reaction sequence of the rTCA cycle is enforced by some combination of the external conditions in the presence of non-selective catalysts that are active in specific types of reactions, the relatives of the rTCA cycle can be identified as the cycles with the most similar sequence of reaction types, even if they have completely different molecular sequences. The reaction type sequence of the rTCA cycle is (1 2 3 5 4 1 2 3 5 6 7). The first carboxylation step "2" transforms pyruvate into oxaloacetate. If this step is skipped, the reaction sequence becomes (1 _ 3 5 4 1 2 3 5 6 7), and a new cycle emerges where pyruvate is reduced to lactate causing further divergence of the molecular sequence from rTCA cycle (see Fig. 5B).

So far in our analysis of the cycles we considered the intrinsic properties that depend on the features of the cycles regardless of their embedding into the supernetwork, such as energetic characteristics of the cycles. Alternatively, the structural properties of the supernetwork can substitute chemical considerations in the choice of the targets. An example of an extrinsic property that depends on the supernetwork structure is node centrality. Node centralities are often discussed in the context of network transport and network robustness[17]. There are multiple definitions of centrality. For example, the lengths of the shortest paths connecting a given molecule to the rest of the network characterize its closeness centrality[18]. One can define the cycle centrality as the lowest centrality of its components. Following this definition, the two cycles shown in Figure 5C are identified as the most central cycles. Both cycles start from glyoxylate; one proceeds to oxaloglycolate and the other to tartrate. Both cycles are shorter than rTCA cycle and include autocatalytic steps.

One possible route to the down-selection in the available cycles could be the selection of a universal carrier of chemical energy exchange. It has been argued that the length of the forward (oxidative) TCA cycle is optimized by evolution to match the discretization of the chemical energy supply in units of adenosine triphosphate (ATP) that is encountered in metabolic systems[19]. This leads us to the question: Would shorter cycles, such as glyoxylate-based cycles from Fig. 5C, have higher relevance in a world where the ATP-based energy carrier may not have been down-selected for? There is another question: How stable is the enzymatic rTCA cycle with respect to the loss of the enzymatic selectivity? Based on the degeneracy of possible cycles encountered in this work, we propose to investigate if modified versions of standard enzymes or promiscuous enzymes exist that facilitate realizations of the substituted/unnatural versions of rTCA cycle, such as the cycles shown in Figs. 5A and 5B. Evidence of the viability of these simple modifications of rTCA cycle will advance our understanding of the origin of the common metabolic core. We would like to emphasize that the proposed supernetwork description captures the diversity of the reaction sequences compatible with a given reactivity model and would encourage the search for any other prebiotic targets and scenarios in this dataset.

Methods

The construction of the chemical supernetwork as introduced in the main text proceeds iteratively. Each iteration includes the following steps:

- The available molecules are validated via quantum chemical calculations (details are explained below); their symbolic representations are produced.
- The symbolic representations of the validated molecules are exhaustively checked against the the reaction types of the model, see transformations in Fig. 1B.
- The reactant molecules are transformed symbolically into the products according to the applicable reaction types.
- The symbolic representations of the new molecules are mapped onto the quantum chemical potential energy surface (PES) for validation.
- The reactions and their products are accepted or rejected on the basis of the quantum chemical validation.
- The list of the available molecules is updated to include the new molecules formed in the accepted reactions.

The symbolic manipulations are performed using cheminformatic tools, such as the simplified molecular-input line entry system (SMILES)[20]. These manipulations include a) transformation of a symbolic representation of a molecule into atomic Cartesian coordinates suitable for quantum chemical calculations, b) transformation of the atomic Cartesian coordinates produced by quantum chemical optimization into symbolic representation of the molecules, and c) symbolic transformation of reactants into products according to the reaction rules.

The semiempirical quantum chemical method PM7[21] as implemented in the MOPAC software [Stewart, J.J.P. MOPAC2012 Stewart Computational Chemistry, Colorado Springs, CO, USA, http://OpenMOPAC.net (2007) Date of access:10/12/2014] was used to carry out the quantum chemical computations. The purpose of the quantum chemical validation was to check if the symbolically generated molecules are viable. Viability is understood as correspondence to a minimum on a quantum chemical potential energy surface that preserves interatomic connectivities of the symbolic representation. The validation, therefore, starts with conversion of the symbolic representation into the Cartesian coordinates of the atoms. This initial structure is optimized to a local minimum on the respective quantum chemical PES. Normal mode analysis is then carried out to ensure that a minimum is reached. Finally, we compute the free Gibbs energy of formation in the gas phase and in aqueous solution within the polarizable continuum medium (PCM) approximation. The solution-phase values are discussed in the text. The free energies are computed under standard conditions at 298 K following the methodology implemented in the MOPAC software package. PCM approximations in the form of the conductor-like screening model (COSMO) are used[22]. The COSMO procedure generates a conducting polygonal surface around the system (ion or molecule), at the van der Waals' distance. By introducing a permittivity-dependent correction factor into the expressions for the screening energy and its gradient, the theory can be extended to finite dielectric constants with only a small error. The standard Gibbs free energy of molecular hydrogen in aqueous solution is set to zero. The described methodology of quantum chemical calculations neither aims at nor implies achieving chemical accuracy of 1 kcal/mol in the energy evaluations. As a form of virtual screening, the presented study relies on robust ranking of the candidates which chosen methodologies of quantum chemical calculations are expected to deliver. The acceptance criteria for the generated molecules and reactions include preserved atomic connectivity between the symbolic notation and optimization results, the local minimum nature of the candidate molecules, and the Gibbs free energies of reactions below 60 kcal/mol. The computations disregard the conformational equilibria so the produced free energies serve as a screening criterion rather then an accurate thermodynamic characteristic. Furthermore, we emulate some of the bias of Refs. 2, 3 with regard to the admissible molecules by checking all generated molecules against the ChemSpider database[23], a modern publicly available equivalent of the Belstein database, and retaining only the molecules with exact matches. The point of using a combination of symbolic and quantum chemistry is clear. It facilitates the develop-

ment of a general methodology of the investigation of prebiotic chemical space that is not limited by the content of the empirical databases. For example, databases are not guaranteed to include thermodynamic data of the reactions of interest.

The chemical supernetwork obtained in this manner incorporates the uncertainty of the model chemistry with respect to the choice of the reaction substrates and reaction sequences. It relates to the single reaction sequence in the same way as a family of curves is related to a single curve. The chemical supernetwork is visualized as a Petri net[24], in which the molecules and the reactions are nodes forming two partitions. Each reaction node and its first-order neighborhood, i.e., the incoming edges with molecular nodes of reactants and the outgoing edges with molecular nodes of products, corresponds to a fragment of a quantum chemical potential energy surface (PES) which is coarse-grained in terms of the Lewis structures describing the local minima (Fig. 2A).

1. Ruiz-Mirazo, K., Briones, C. & de la Escosura, A. Prebiotic systems chemistry: new perspectives for the origins of life. *Chem. Rev.* **114**, 285–366 (2014).
2. Morowitz, H. J., Kostelnik, J. D., Yang, J. & Cody, G. D. The origin of intermediary metabolism. *Proc. Natl. Acad. Sci. USA* **97**, 7704–7708 (2000).
3. Smith, E. & Morowitz, H. J. Universality in intermediary metabolism. *Proc. Natl. Acad. Sci. USA* **101**, 13168–13173 (2004).
4. Buchanan, B. B. & Arnon, D. I. A Reverse Krebs cycle in photosynthesis: consensus at last. *Photosynth. Res.* **24**, 47–53 (1990).
5. Hügler, M. & Sievert, S. M. Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Annu. Rev. Marine Sci.* **3**, 261–289 (2011).
6. Orgel, L. E. Self-organizing biochemical cycles. *Proc. Natl. Acad. Sci. USA* **97**, 12503–12507 (2000).
7. Orgel, L. E. The implausibility of metabolic cycles on the prebiotic Earth. *PLoS Biology* **6**, 5–13 (2008).
8. Ross, D. S. The viability of a nonenzymatic reductive citric acid cycle - kinetics and thermochemistry. *Orig. Life Evol. Biosph.* **37**, 61–65 (2007).
9. Zhang, X. V. & Martin, S. T. Driving parts of Krebs cycle in reverse through mineral photochemistry. *J. Am. Chem. Soc.* **128**, 16032–16033 (2006).
10. Saladino, R. *et al.* Photochemical synthesis of citric acid cycle intermediates based on titanium dioxide. *Astrobiology* **11**, 815–24 (2011).
11. Cooper, G., Reed, C., Nguyen, D., Carter, M. & Wang, Y. Detection and formation scenario of citric acid, pyruvic acid, and other possible metabolism precursors in carbonaceous meteorites. *Proc. Natl. Acad. Sci. USA* **108**, 14015–14020 (2011).
12. Sagi, V. N., Punna, V., Hu, F., Meher, G. & Krishnamurthy, R. Exploratory experiments on the chemistry of the "glyoxylate scenario": formation of ketosugars from dihydroxyfumarate. *J. Am. Chem. Soc.* **134**, 3577–3589 (2012).
13. Guzman, M. I. & Martin, S. M. Photo-production of lactate from glyoxylate: how minerals can facilitate energy storage in a prebiotic world. *Chem. Comm.* **46**, 2265–2267 (2010).
14. Butch, C. *et al.* Production of tartrates by cyanide-mediated dimerization of glyoxylate: a potential abiotic pathway to the citric acid cycle. *J. Am. Chem. Soc.* **135**, 13440–13445 (2013).
15. Russi, T., Packard, A. & Frenklach, M. Uncertainty quantification: Making predictions of complex reaction systems reliable. *Chem. Phys. Lett.* **499**, 1–8 (2010).
16. Crick, F. H. C. The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
17. Woolley-Meza, O. *et al.* Complexity in human transportation networks: a comparative analysis of worldwide air transportation and global cargo-ship movements. *Eur. Phys. J. B* **84**, 589–600 (2011).
18. Sabidussi, G. Centrality index of a graph. *Psychometrika* **31**, 581–603 (1966).
19. Bar-Even, A., Noor, E., Lewis, N. E. & Milo, R. Design and analysis of synthetic carbon fixation pathways. *Proc. Natl. Acad. Sci. USA* **107**, 8889–8894 (2010).
20. Weininger, D. J. SMILES, a chemical language and information system. *Chem. Inf. Model.* **28**, 31–36 (1966).
21. Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **19**, 1–32 (2013).
22. Klamt, A. & Schüürmann, G. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Transactions* **2**, 799–805 (1993).
23. Pence, H. E. & Williams, A. ChemSpider: an online chemical information resource. *J. Chem. Educ.* **87**, 1123–1124 (2010).
24. Murata, T. Petri nets: properties, analysis, and applications. *Proc. IEEE* **77**, 541–580 (1989).

Acknowledgments

This work was supported by a grant from the Simons Foundation (SCOL 291937, D.Y.Z.) as well as the National Science Foundation Cyberdiscovery Initiative Type II (CDI2) grant number OIA-1125087 (D.R. and A.A.-G.).

Author contributions

D.Y.Z. proposed the project and carried out computations. D.Y.Z. and D.R. developed the methodology of the study and analyzed the data, A.A.-G. supervised the project. D.Y.Z., D.R. and A.A.-G. wrote the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zubarev, D.Y., Rappoport, D. & Aspuru-Guzik, A. Uncertainty of Prebiotic Scenarios: The Case of the Non-Enzymatic Reverse Tricarboxylic Acid Cycle. *Sci. Rep.* **5**, 8009; DOI:10.1038/srep08009 (2015).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/