

Artificial Chemistry Performed in an Agglomeration of Droplets with Restricted Molecule Transfer [★]

Johannes Josef Schneider^{1,2}[0000-0002-6921-9049], Alessia Faggian³[0000-0002-3799-9722], William David Jamieson²[0000-0001-8260-5211], Mathias Sebastian Weyland¹[0000-0001-6389-0029], Jin Li²[0000-0002-4672-6806], Oliver Castell²[0000-0002-6059-8062], Hans-Georg Matuttis⁴, David Anthony Barrow²[0000-0003-2096-7262], Aitor Patiño Diaz³[0000-0001-5692-3822], Lorena Cebolla Sanahuja³, Silvia Holler³[0000-0002-2572-8677], Federica Casiraghi³[0000-0001-8233-049X], Martin Michael Hanczyc^{3,5}[0000-0002-5453-2139], Dandolo Flumini¹[0000-0003-3699-4871], Peter Eggenberger Hotz¹, and Rudolf Marcel Fuchsli^{1,6}[0000-0002-9296-5314]

¹ Institute of Applied Mathematics and Physics, School of Engineering, Zurich University of Applied Sciences, Technikumstr. 9, 8401 Winterthur, Switzerland
johannesjosefschneider@googlemail.com, {scnj,weyl,flum,egg,escl,furu}@zhaw.ch
<https://www.zhaw.ch/en/about-us/person/scnj/>

² School of Engineering, Cardiff University, Queen's Buildings, 14-17 The Parade, Cardiff CF24 3AA, Wales, United Kingdom
{jamiesonw,LiJ40,Castell10,Barrow}@cardiff.ac.uk

³ Laboratory for Artificial Biology, Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, 38123, Trento, Italy
{alessia.faggian,aitor.patino,lorena.cebolla,silvia.holler,federica.casiraghi,martin.hanczyc}@unitn.it

⁴ Department of Mechanical Engineering and Intelligent Systems, The University of Electrocommunications, Chofu Chofugaoka 1-5-1, Tokyo 182-8585, Japan
hg@mce.uec.ac.jp

⁵ Chemical and Biological Engineering, University of New Mexico, MSC01 1120, Albuquerque, NM, 87131-0001, USA

⁶ European Centre for Living Technology, S.Marco 2940, 30124 Venice, Italy

Abstract. Within the scope of the European Horizon 2020 project *ACDC – Artificial Cells with Distributed Cores to Decipher Protein Function*, we aim at the development of a chemical compiler governing the three-dimensional arrangement of droplets, which are filled with various chemicals. Neighboring droplets form bilayers with pores which allow chemicals to move from one droplet to its neighbors. With an appropriate three-dimensional configuration of droplets, we can thus enable gradual biochemical reaction schemes for various purposes, e.g., for the production of macromolecules for pharmaceutical purposes. In this paper, we

[★] This work has been partially financially supported by the European Horizon 2020 project *ACDC – Artificial Cells with Distributed Cores to Decipher Protein Function* under project number 824060.

demonstrate with artificial chemistry simulations that the ACDC technology is excellently suitable to maximize the yield of desired reaction products or to minimize the relative output of unwanted side products.

Keywords: artificial chemistry · agglomeration · droplet · preference

1 Introduction

1.1 Background

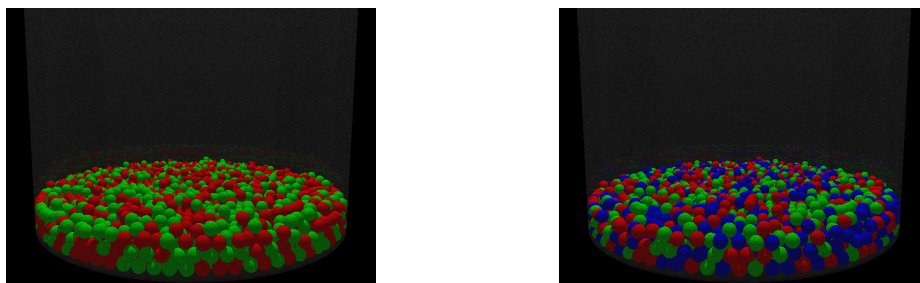


Fig. 1. Agglomeration of 3,000 spherical droplets with radius $30\mu\text{m}$ at the bottom of a cylinder, created in a computer simulation as described in [9]. The configurations in both pictures are identical with respect to the geometric locations of the droplets, but the left graphics shows a bipartite system with two particle types, marked as red and green, whereas the right graphics shows a tripartite system with three particle types, marked as red, green, and blue. These two configurations serve as the common underlying geometric structure and, with respect to particle types, as starting points for our artificial chemistry simulations.

In organic chemistry, one usually deals with molecules containing functional groups, like the hydroxyl $-\text{OH}$ group of alcohols, the carbonyl $-\text{C}=\text{O}$ group of ketones and aldehydes, and the carboxyl $-\text{COOH}$ group of organic acids. These and other functional groups allow specific reactions, e.g., an alcohol and a carboxylic acid can form an ester, while splitting off a water molecule. Besides these well-known natural functional groups, additionally corresponding pairs of artificial functional groups have been created in the last decades in the field of click chemistry [5,3] to better govern reaction processes. However, the number of different active groups which can be used simultaneously in a reaction environment as intended is usually rather limited [13], which limits this approach.

Therefore, we apply another approach: Instead of having one well-stirred pot containing all educts for the chemical reaction process, we use droplets within a hull of amphiphilic molecules like phospholipids. The droplets are filled with only one chemical each at the beginning and are put in a cylinder where they

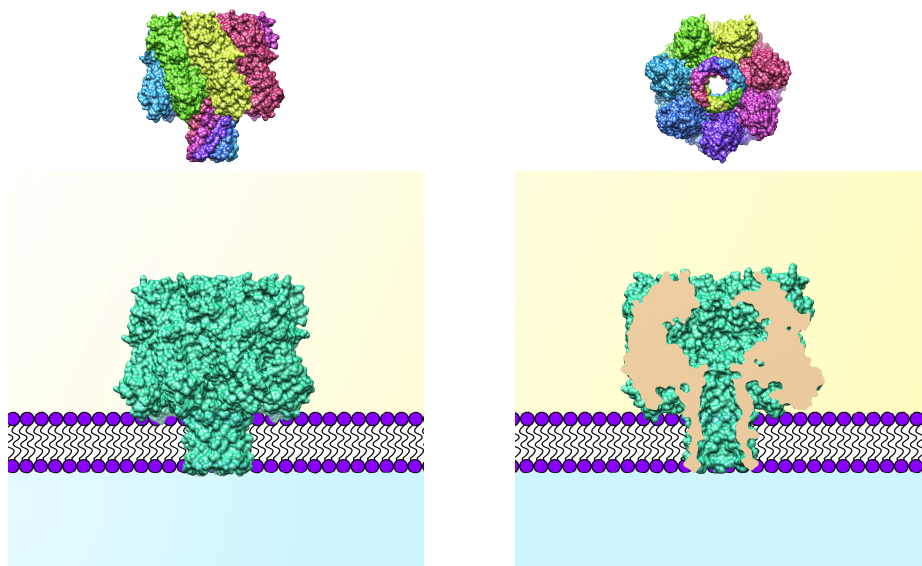


Fig. 2. Alpha hemolysin (aHL) pore: The top row shows the aHL, which is comprised of seven macromolecules, from the side view (left), displaying its trunk, and from the bottom up view (right), displaying the narrow channel in its center [12]. The bottom row reveals how the aHL sticks its trunk through the bilayer between two droplets (left) and presents in detail a cut through the pore formed (right).

sink and agglomerate at the bottom. Such an agglomeration is shown in Fig. 1. In the computer simulation, 3,000 spherical droplets were first placed randomly in a cylinder and sink to the bottom under the influence of gravity reduced by the buoyant force, Stokes friction, added mass effect, small random velocity changes, and quasi-elastic collisions among each other and with the walls of the cylinder [11,9]. Droplets touching each other can form a small bilayer and are then able to exchange small molecules through pores in these bilayers, as shown at the example of an alpha hemolysin (aHL) pore in Fig. 2. The creation of such connections can also be restricted so that neighboring pairs can only form connections if additional conditions are fulfilled. Such additional conditions can be realized by e.g. placing the seven constituent molecules of aHL only in some of the droplets. Covering the hulls of the droplets with single strands of DNA, as shown in Fig. 3, is another way of imposing such restrictions. Only if neighboring droplets are covered with complementary strands of DNA, a connection for exchanging molecules can be formed. Such restrictions offer a wide variety of altering the time evolution of gradual chemical reaction schemes, which will be useful, as we will show later. Furthermore, this overall approach with an agglomeration of droplets with bilayers which allow the exchange of molecules mimicks

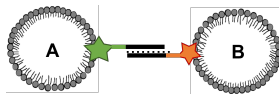
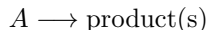


Fig. 3. Sketch of a pair of oil-filled droplets in water, to which complementary strands of ssDNA oligonucleotides are attached: The surfaces of the droplets are composed by single-tail surfactant molecules like lipids with a hydrophilic head on the outside and a hydrophobic tail on the inside, thus forming a boundary for the oil-in-water droplet. By adding some single-strand DNA to the surface of a droplet, it can be ensured that only desired connections to specific other droplets with just the complementary single-strand DNA can be formed. Please note that the connection of the droplets in this picture is overenlarged in relation to the size of the droplets. In realiter, the droplets have a radius of $1 - 50\mu\text{m}$, whereas a base pair of a nucleic acid is around 0.34nm in length [1], such that the sticks of connecting DNA strands are roughly 5nm long.

living organisms with their spatial structuring. The droplets correspond to the cells and their hulls to the membranes. One can also use a chemist’s language and call the droplets the containers and the exchange of molecules transport processes. Mathematically speaking, the agglomeration forms a graph with the droplets being the nodes and the connections being the edges.

1.2 Gillespie Algorithm

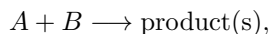
While reaction kinetics is often modeled with ordinary differential equations, in solutions in organic chemistry the number of molecules is too small to justify the assumption of continuous concentrations. Therefore, we use Gillespie’s stochastic simulation method [4], in a way called the ‘direct method’ [2], which we want to briefly describe here: Gillespie’s stochastic simulation method deals with the scenario in which there are various chemical reactions possible in the system. For each possible reaction i , a so-called propensity $p(i)$ has to be calculated. The simplest possible reaction is the unimolecular reaction with one educt A only:



In order to calculate the propensity of this reaction i , one needs the number $N(A)$ of molecules of type A and its reaction constant $k(i)$. Then the propensity of this unimolecular reaction is given by

$$p(i) = N(A) \times k(i). \quad (1)$$

For a bimolecular reaction j with two different educts A and B , i.e.



the propensity is given by

$$p(j) = N(A) \times N(B) \times k(j)/V. \quad (2)$$

Here the number of possible combinations of an A and a B molecule is multiplied with the reaction constant $k(j)$, which has to be renormalized by the volume V of the container [2]: The larger the volume, the less likely it is that an A molecule comes near a B molecule. To simplify matters, we set all $k(i) \equiv k = 1$ and, as we work with a monodisperse system in which each droplet has the same volume $V(i) \equiv V$, we set $V = 1$. According to [2], the transport of a molecule can be treated like a unimolecular reaction. In the next step, one needs to sum up all n propensities into an overall propensity P , $P = \sum_{i=1}^n p(i)$. Then one needs to choose two independent random numbers r and z , which are uniformly chosen from the interval $(0; 1)$. From z , one derives an exponentially distributed random number e by setting $e = -\ln(z)$. The time interval between the previous reaction and the current reaction is given by e/P . While z and e are used to determine the time in which the new reaction takes place, the other random number r is used to determine which of the reactions is to be chosen. That reaction i is chosen for which the condition

$$\sum_{j=1}^{i-1} p(j) < r \times P \leq \sum_{j=1}^i p(j) \quad (3)$$

holds. This way, a series of reactions can be created with the Gillespie algorithm, taking care of the correct time scale and the different propensities of the various reactions. The algorithm stops if the overall propensity P is zero, meaning that there is no further reaction possible, or if a predefined maximum amount of time has evolved [2].

This Gillespie algorithm can be easily extended to a system like ours with various containers, so that all possible reactions within each container and all possible transport processes between each pair of containers are represented in the $p(i)$, which are summed up in P . With the random number r , one selects the respective $p(i)$ to determine not only the reaction but also the container or the transport process between a pair of containers.

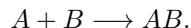
In the next sections, we study various scenarios for connections between the droplets in the spatial arrangement shown in Fig. 1 for simple reaction schemes.

2 Reaction with two educts

Table 1. Properties of the networks, as described in scenarios 1 and 2 in Sec. 2

	scenario 1	scenario 2
number of edges	8983	4495
maximum degree of a node	12	10
mean degree of a node	5.98	2.99
number of clusters	1	114
maximum cluster size	3000	2876

The simplest bimolecular reaction is



This reaction, which shall be studied in this section, has two educt molecules A and B and one product molecule $A - B$, shortly written as AB . In this section, we do not allow any further reactions. Additionally to this bimolecular reaction, we have transport processes of the three different molecules A , B , and AB . The overall number of transport processes for a droplet i is thus three times the number of droplets droplet i is attached to. In order to get a well-stirred agglomeration of droplets, we might want to create connections between every pair of neighboring droplets. This scenario, to which we want to refer to as scenario 1 in this section, should lead to a fast stirring and a fast reaction process. But one might wonder whether it would not be better to allow only connections between droplets which initially only contain A molecules and those which initially only contain B molecules. Then the molecules would not need to jump first to other molecules of their own kind but can directly move into those containers which initially contain only their desired reaction partners for the reaction process $A + B \rightarrow AB$. Thus, this scenario 2 might lead to an even faster reaction process.

Using the droplet agglomeration comprised of 1,500 droplets filled with 1,000 molecules A and 1,500 droplets filled with 1,000 molecules B in Fig. 1, we create a network each for these two scenarios and perform a network analysis, with the most important results given in Tab. 1. We find that at least one droplet has indeed the maximum possible number of neighboring droplets in scenario 1, which equals the kissing number in three dimensions [6,10], but the mean number of droplets a droplet is attached to is only roughly 6 in scenario 1 and 3 in scenario 2. In scenario 1, all 3,000 droplets form one cluster, whereas there is a large dominating cluster containing most of the droplets, some unconnected droplets, and a few small separate clusters in scenario 2. (Please note that for other agglomerations, also for scenario 1, the largest cluster might not contain all droplets.)

In the next step, we have a look at the time evolution of the reaction $A+B \longrightarrow AB$ in these two scenarios, which is shown in Fig. 4. For comparison, we also show the reaction dynamics, if we would throw all molecules into a single pot with a volume of 3,000 times the volume of one droplet. In order to study the dynamics also at short time scales, all diagrams are drawn with a logarithmic time axis. We find sigmoidal decreases of $N(A)$ and $N(B)$ (the curves for the time evolutions of $N(A)$ and $N(B)$ coincide), while the number $N(AB)$ of the product molecule increases correspondingly. The pot scenario depicts the fastest dynamics by far. The more elaborate scenario 2, which we hoped would generate a faster dynamics than scenario 1, turns out to be equally fast only at short time scales, for medium time scales it slows down. And even worse, as some of the droplets are not connected to the largest cluster, not all educts are able to react, such that the yield is also smaller when using this scenario. Summarizing, we have to state that according to the results presented so far, the ACDC technology seems to

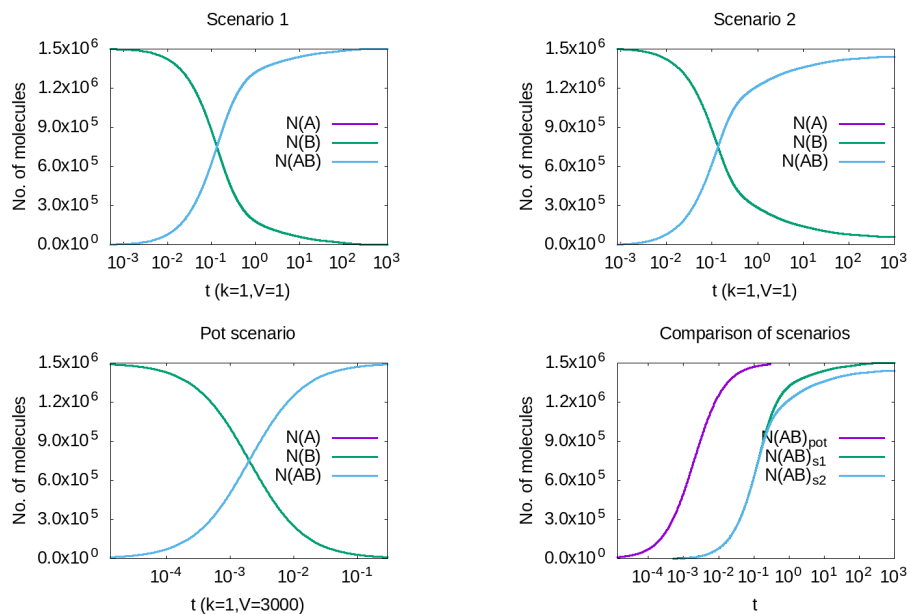
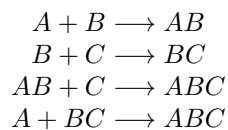


Fig. 4. Results for the time evolution of the reaction process $A+B \rightarrow AB$, for scenario 1 (top left), scenario 2 (top right), and the pot scenario (bottom left), as described in the text. The picture at the bottom right replots the increases of the numbers of AB molecules for the three scenarios for a better comparison.

be disadvantageous compared to a well stirred pot and trying to improve the method with a seemingly clever ansatz even leads to a significant deterioration.

3 Reaction with three educts and one desired product

The next more complex example for a reaction scheme would of course involve three different types of educts A , B , and C to create a molecule $A-B-C$ (short ABC) via a gradual reaction scheme in two steps. Hereby we allow four reactions as follows:



As a starting point, we use the right graphic in Fig. 1, with 1,000 red droplets initially filled with 1,000 molecules of type A , 1,000 green droplets initially filled with 1,000 molecules of type B , and 1,000 blue droplets initially filled with 1,000

Table 2. Results for the networks obtained for the four scenarios as described in Sec. 3. The same observables as in Tab. 1 are presented.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
number of edges	8983	6013	3906	6876
maximum degree	12	12	12	12
mean degree	5.98	4.00	2.60	4.58
number of clusters	1	22	270	14
maximum cluster size	3000	2979	2680	2986

molecules of type C . Also here we investigate various scenarios: In scenario 1, each droplet is connected with each of its neighbors as in the previous section. For scenario 2, we want to generalize the scenario 2 of the last section. Thus, in scenario 2, each droplet is connected to neighbors with a different color, i.e., red-red, green-green, and blue-blue connections are not allowed, but all other connections are allowed. But we could also adopt scenario 2 of the last section in another way by trying to steer the reaction process in the right direction. Thus, in scenario 3, we only allow connections red and green droplets and connections between green and blue droplets. However, in this scenario 3, the large cluster might be split in many small clusters. In order to overcome this disadvantage, we define a scenario 4, which contains the same connections as in scenario 3, but also red-red, green-green, and blue-blue connections. Thus, in scenario 4, only red-blue connections are forbidden. Again we perform a network analysis and show the most important results in Tab. 2. We find that the seemingly cleverest scenario 3 contains the smallest number of edges, but also the largest number of smaller sized clusters.

Figure 5 displays the time evolution of the numbers of molecules for the three educts A , B , and C , for the intermediary products AB and BC , and for the final product ABC , for the scenarios described above and also, for comparison, for the pot scenario again. We generally find a sigmoidal decrease for all three educts, with the decrease of educt B being much faster than those of educts A and C , which coincide. This deviation can be easily explained, as molecule B is part of both reaction processes leading to intermediary products. The two intermediary products show an almost identical rise and decrease again. When comparing the sigmoidal increase of the number of ABC molecules for the various scenarios, we find that of course again the well-stirred pot is much faster than any of the ACDC technology scenarios. Furthermore, we find for short time scales that scenarios 1 and 2 as well as scenarios 3 and 4 exhibit the same dynamics, while at medium time scales, scenario 4 approaches the dynamics of scenarios 1 and 2. The optimum yield is of course only obtained with scenario 1, but it takes a long time for the last educts to move around in the labyrinth of the agglomeration until a partner molecule for a reaction can be found. The worst yield is obtained with scenario 3, which again was originally designed to help and accelerate the gradual reaction processes. Summarizing, also here we have to state that the application of the ACDC technology seems to be disadvantageous when compared to a well-

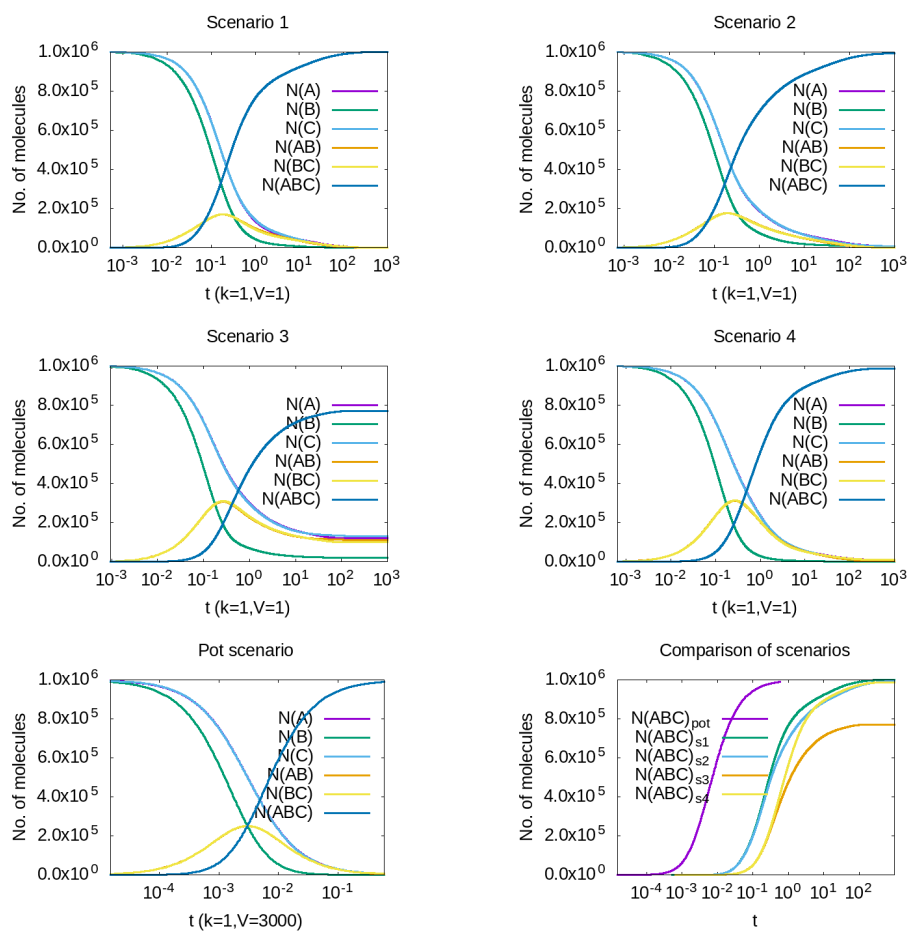
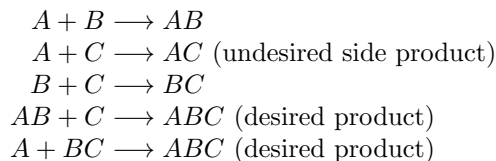


Fig. 5. Results for the time evolution of the gradual chemical reaction scheme for the scenarios described in Sec. 3. The overall numbers of the educt molecules A , B , and C , of the intermediary products AB and BC , and of the final product ABC exhibit a rich behavior. Like in Fig. 4, the graphic at the bottom right displays a comparison of the dynamics of the various scenarios by replotting the sigmoidal increase of the overall number of product molecules.

stirred pot and the attempt to improve it leads to a significant deterioration again.

4 Reaction scheme with three educts, one desired product, and one undesired side product

In the next step, we extend slightly the reaction scheme presented in the last section with a further reaction, leading to an undesired side product AC . Thus, we have:



Only these five reactions shall be allowed, other reactions are not possible. Furthermore, we use the scenarios as defined in the previous section again. (As we use the same scenarios, the results presented in Tab. 2 apply also in this section.)

Table 3. Final numbers of molecules for the desired product ABC and for the undesired side product AC and their ratio for the scenarios as described in Sec. 4.

	$N(ABC)$	$N(AC)$	$N(AC)/N(ABC)$
Scenario 1	408769	357343	0.874
Scenario 2	406865	355197	0.873
Scenario 3	640700	71112	0.111
Scenario 4	798561	96804	0.121
Pot scenario	399811	367796	0.920

Figure 6 presents the results for the reaction scheme defined in this section. Again we find a sigmoidal decrease of the educts (the curves for A and C coincide again), an intermediate peak of the intermediary products AB and BC (their curves coincide, too), and an sigmoidal increase both for ABC and AC . We find that the introduction of an undesired side product changes the final outcomes for the scenarios dramatically. Still, the pot scenario leads to the fastest dynamics. But all scenarios using the ACDC technology lead to a larger yield of the desired product ABC . When we aim at maximizing the number of ABC molecules, the best results are obtained with scenario 4, followed by scenario 3. The exact final values for the numbers of molecules of the desired product ABC and the undesired side product AC are provided in Tab. 3, together with the ratio between them, as instead of maximizing $N(ABC)$, one might want to aim at minimizing the ratio between the undesired side product and the desired product. While scenario 4 leads to the by far largest yield of the desired product ABC , scenario 3 provides a slightly better ratio. The question which of these two scenarios shall be chosen depends on the question how important the condition to minimize the ratio is.

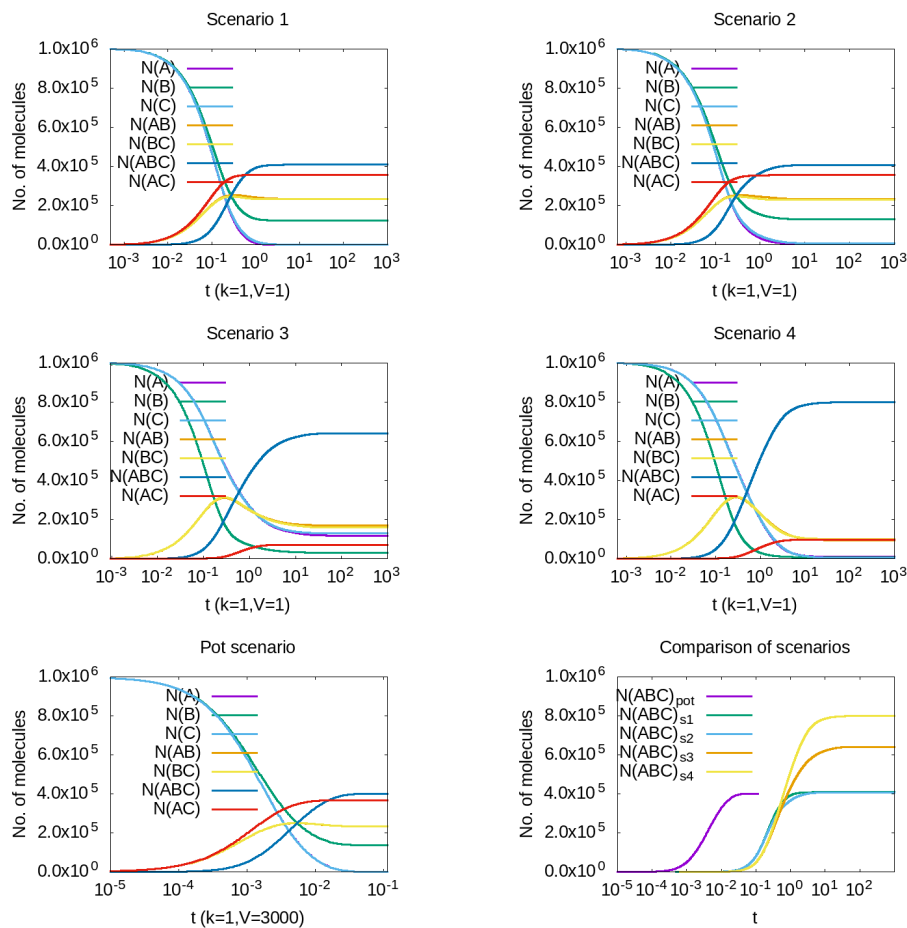


Fig. 6. Results for the reaction processes defined in Sec. 4. The results are shown in the same way as in Fig. 5 for the reaction processes of Sec. 3, but additionally with curves for the side product AC .

5 Conclusion and Outlook

In this paper, we applied some small toy instances of artificial chemistries to a network of droplets, which are initially filled with only one chemical each. In comparison to a well-stirred pot, in which one would usually put all the ingredients for the desired gradual reaction scheme, such an agglomeration of droplets allows for some additional steering possibilities for the reaction scheme by restricting the types of connections in the network via which molecules can move from one droplet to another one. While this network approach leads to a much slower dynamics than the well-stirred pot, one can show that an appropriate

choice of connections can lead to much better results, if undesired side products occur in the reaction scheme. Summarizing the results for these simple examples we can state that the ACDC technology is indeed a highly useful tool for gradual reaction schemes with at least two steps and at least one undesired side product. An appropriate choice of connections helps maximizing the yield or minimizing the ratio between the amount of undesired side products and the amount of desired end products – two aims which are not entirely identical. We intend to apply this approach to more complex reaction schemes [13] and to study the development of increasingly complex molecules both in pharmaceuticals as in [7] and in the area of questioning the origins of life [8].

References

1. Annunziato, A.T.: DNA Packaging: Nucleosomes and Chromatin. *Nature Education* **1**, 26 (2008)
2. Cieslak, M., Prusinkiewicz, P.: Gillespie-lindenmayer systems for stochastic simulation of morphogenesis. in *silico Plants* **1**, diz009 (2019)
3. Devaraj, N.K., Finn, M.: Introduction: Click chemistry. *Chemical Reviews* **121**, 6697–6698 (2021)
4. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* **22**, 403–434 (1976)
5. Kolb, H.C., Finn, M., Sharpless, K.B.: Click chemistry: Diverse chemical function from a few good reactions. *Angew. Chem., Int. Ed.* **40**, 2004–2021 (2001)
6. Kucherenko, S., Belotti, P., Liberti, L., Maculan, N.: New formulations for the kissing number problem. *Discrete Applied Mathematics* **155**, 1837–1841 (2007)
7. Marshall, S.M., Mathis, C., Carrick, E., Keenan, G., Cooper, G.J.T., Graham, H., Craven, M., Gromski, P.S., Moore, D.G., Walker, S.I., Cronin, L.: Identifying molecules as biosignatures with assembly theory and mass spectrometry. *Nature Communications* **12**, 3033 (2021)
8. Oparin, A.: *The Origin of Life on the Earth*, 3rd ed. Academic Press, New York (1957)
9. Schneider, J.J., Faggian, A., Patiño Diaz, A., Li, J., Holler, S., Casiraghi, F., Cebolla Sanahuja, L., Matuttis, H.G., Hanczyc, M.M., Barrow, D.A., Weyland, M.S., Flumini, D., Eggenberger Hotz, P., Fuchsli, R.M.: Network creation during agglomeration processes of monodisperse and polydisperse systems of droplets (2022), submitted to *microTAS 2022*
10. Schneider, J.J., Barrow, D.A., Jin, L., Weyland, M.S., Flumini, D., Eggenberger Hotz, P., Fuchsli, R.M.: Geometric restrictions to the agglomeration of spherical particles (2021), accepted for publication in *Wivace conference proceedings*, *Wivace 2021*
11. Schneider, J.J., Faggian, A., Holler, S., Casiraghi, F., Li, J., Cebolla Sanahuja, L., Matuttis, H.G., Hanczyc, M.M., Barrow, D.A., Weyland, M.S., Flumini, D., Eggenberger Hotz, P., Fuchsli, R.M.: Influence of the geometry on the agglomeration of a polydisperse binary system of spherical particles. vol. *ALIFE 2021: The 2021 Conference on Artificial Life (2021)*. https://doi.org/10.1162/isal_a.00392
12. Song, L., Hobaugh, M.R., Shustak, C., Cheley, S., Bayley, H., Gouaux, J.E.: Structure of staphylococcal α -hemolysin, a heptameric transmembrane pore.

- Science **274**(5294), 1859–1865 (1996), structure was uploaded by the authors on <https://www.rcsb.org/structure/7ahl> and was visualized using Chimera <https://www.cgl.ucsf.edu/chimera>
13. Weyland, M.S., Fellermann, H., Hadorn, M., Sorek, D., Lancet, D., Rasmussen, S., Fuchsli, R.M.: The matchit automaton: Exploiting compartmentalization for the synthesis of branched polymers. *Computational and Mathematical Methods in Medicine* **2013**, 467428 (2013)