Michał Bajczyk

Biomimetic architectures, auto-replication and signal transduction in the systems of organic reactions

Doctoral research realized in group XI
Institute of Organic Chemistry
Polish Academy of Sciences
in years 2014-2018
under supervision of prof Bartosz Grzybowski

Warsaw 2019

This doctoral research was a part of the "Development of Computational Chemical Linguistics and its Applications to the Efficient Planning of Multistep Chemical Syntheses" project, funded by the National Science Centre (SYMFONIA2014/12/W/ST5/00592).

NATIONAL SCIENCE CENTRE
POLAND

I would like to express my deep gratitude to the following people:

Prof. Robert Hołyst, for all the time that he has so generously devoted to sharing his knowledge and helping me improve as a scientist, as well as for his enthusiastic encouragement that allowed me to carry my research through.

Dr. Przemysław Szafrański, my long-time friend, for patiently supporting me for the entire duration of my Ph.D. studies with his thoughtful insights and useful advice.

My younger sister, Natasza, for housing me whilst I was writing this thesis.

Thesis abstract

The presented thesis describes the doctoral research regarding chemical systems, conducted by means of the computer-assisted methods developed by the Grzybowski Group. Said methods comprise the two software platforms, i.e., Cyclorg, allowing for discovery and design of cyclic reaction networks, as well as Kinetix, allowing for numerical analysis of the molecular interactions therein.

In result of the Cyclorg-based research, a series of chemical cycles was identified from the output of published organic reactions, thereby proving that independent chemists working at different times have jointly – yet often unknowingly – created examples of cyclic reaction sequences. Cycles are a prominent feature of metabolism as well as cellular signaling, thereby attracting the interest from systems chemistry and biology.

The Kinetix-based research has shown that chemical reactions, as well as entire networks thereof, operate similarly to low-pass filters when responding to the reactant concentration oscillations introduced by external means. The study of the realistic kinase cascade model allowed for discovery of an emergent property thereof, which allows for transforming time-varying signals into continuous ones, i.e., the higher the frequency of the introduced oscillatory signal, the higher the average strength of the received continuous signal.

The dissertation is divided into four parts comprising (1) the introduction regarding systems chemistry, (bio)chemical cycles and the cascades thereof as well as the emergent properties relevant to the presented work, (2) the section describing the Cyclorg-based research and the examples of identified chemical cycles, (3) the section describing step-by-step the scientific process regarding the analysis of signal propagation in (bio)chemical systems by means of Kinetix software, and (4) summary of the results, which includes the equations of signal propagation, established over the course of the doctoral research.

Streszczenie pracy

Niniejsza praca opisuje badania dotyczące układów chemicznych przeprowadzone z użyciem metod wspomaganych komputerowo, opracowanych przez grupę prof. Grzybowskiego. Wspomniane metody obejmują dwie platformy programowe: Cyclorg, pozwalający odkrywać i projektować cykliczne układy chemiczne, oraz Kinetix, pozwalający badać kinetykę oddziaływań wewnątrz tychże układów metodami numerycznymi.

W wyniku badań opartych na platformie Cyclorg zidentyfikowano szereg cykli chemicznych wśród zbiorów opublikowanych reakcji organicznych, tym samym udowadniając, że samodzielni chemicy pracujący w różnych czasach wspólnie – acz niekoniecznie świadomie – stworzyli przykłady cyklicznych sekwencji reakcji chemicznych. Cykle są istotnym elementem metabolizmu oraz sygnalizacji komórkowej, tym samym ciesząc się zainteresowaniem w chemii i biologii układów.

Badania oparte na platformie Kinetix pokazały, że reakcje chemiczne, a także całe ich układy, działają podobnie do filtrów dolnoprzepustowych kiedy oddziałują z oscylacjami stężeń wprowadzonymi z zewnątrz. Badania nad realistycznym modelem kaskady kinaz pozwoliły na odkrycie jej własności emergentnej, pozwalającej na przekształcenie sygnału zmiennego w stały, tzn. im wyższa częstotliwość wprowadzonego oscylującego sygnału, tym większa średnia wartość otrzymanego stałego sygnału.

Przedstawiona praca doktorska jest podzielona na cztery główne części zawierające (1) wstęp do chemii układów, cykli (bio)chemicznych i ich kaskad a także własności emergentnych istotnych dla tej pracy, (2) rozdział opisujący badania oparte na platformie Cyclorg oraz przykłady zidentyfikowanych cykli, (3) rozdział opisujący krok po kroku proces analizy przenoszenia sygnału przez układy (bio)chemiczne za pomocą platformy Kinetix, oraz (4) streszczenie wyników zawierające m.in. równania opisujące przenoszenie sygnałów opracowane w trakcie opisanych badań.

Table of Contents:

1. Aim of this doctoral research

To explore systems chemistry by means of computational software from the Grzybowski Group – Cyclorg and Kinetix.

2. Introduction to the thesis

The work is divided into two parts focusing on different, but conceptually related, aspects of reaction systems.

The first part describes my work on computer-assisted methods for the discovery of new biomimetic chemical-organic cycles and conceptualization of more predictive methods of their discovery.

The second part describes models and numerical calculations aimed at gaining basic understanding of signal propagation in chemical reaction networks and of the ability of such networks to transfer time-varying chemical signals.

During my Ph.D. studies, these two projects relied heavily on the separate pieces of the software created within the Grzybowski Group. My main task was to use these platforms as tools assisting the process of scientific discovery and allowing for verification of various hypotheses related to systems chemistry.

In the case of the first software platform – called Cyclorg – I participated in its creation from the very beginning, although all the coding and calculations were done by my colleague mathematician, Dr. Piotr Dittwald.

Written in Python, Cyclorg is a repository of cycle candidates algorithmically identified within the so-called Network of Organic Chemistry (NOC)[1] – that is, within a set of organic-chemical reactions reported in literature and patents from 1789 to 2011. Transformation of the NOC into a directed graph (where substrates and products are nodes connected by reactions) allowed for its algorithmic navigation, ultimately assembling published reactions into cycles. The searches were performed by means of the standard depth-limited depth-first search (DLDFS) algorithm,[2] using NetworkX library.[3] Multiple searches were performed at the various stages of the project, each resulting in a unique collection of cycles – Cyclorg comprises the three most comprehensive collections, accessible through “database” sub-menu.

When validating Cyclorg-based predictions, my conceptual input was, inter alia, to design filters that would (i) eliminate chemically nonsensical cycles and (ii) allow for specifying search criteria narrowing down the output to cycles of only specific structures and "properties". Those features were implemented into the code perceptibly improving the user experience and chemical relevance.[4] In order to further reinforce the validity of Cyclorg-based results, I have identified and interpreted a wide assortment of plausible and/or interesting cycle candidates.

My second sub-project used the so-called Kinetix software, developed in our group by Dr. Rafał Roszak. Kinetix is a web application, written in JavaScript using the Meteor framework,[5] which simulates kinetics of entire chemical networks. It does so by solving sets of ordinary differential equations by means of the Crank-Nicolson method with an adaptive time step.[6,7]

Upon joining the project, I designed a new functionality of time-dependent concentrations and used it to analyze how periodic changes in substrate concentrations affect the kinetics of the entire reaction networks. At the later stages of the project, I also coded some of the new features and UI fixes myself (although most of the coding was still done by Dr. Roszak), making the final version of Kinetix more user-friendly.[8]

During the Kinetix project, I worked in close collaboration with Prof. Robert Hołyst from the Institute of Physical Chemistry of the Polish Academy of Sciences, who was remarkably helpful, especially with the mathematical analyses, but also a great mentor.

3. Theoretical introduction

3.1. Systems chemistry

Systems chemistry is a young and rapidly developing field of study that deals with emergent properties of entire reaction networks (i.e., complex mixtures of interacting molecules), such as self-replication, self-organization, autoregulation, breaking of chiral symmetry, emergence and transduction of time-varying signals (i.e., concentration oscillations).[9,10] Such properties cannot be attributed to any single component of the system – instead, they are the result of interactions between the system components.[11,12] In order to study them, the field combines classical knowledge of chemistry with the insights and analogies from biology and theoretical framework of computer science of complex systems.[9]

The first mentions of systems chemistry as a discipline date back to 2005, when the term appeared twice in the same week – first in the title of a scientific paper,[13] then in the title of the post-conference workshop of Chembiogenesis meeting in Venice.[9] A series of related discoveries was published afterwards[14] and a new peer-reviewed journal entitled "Systems Chemistry" was founded in 2010.[15] Although the journal was discontinued five years later,[16] the systems chemistry research is still being pursued widely and published in high-impact journals,[17,18] including my own doctoral work described in this thesis.[19,20] The interest in the new discipline is reflected, for instance, in the establishment of the dedicated Centre for Systems Chemistry in Groningen, 2008, which played an important role in securing a multi-million-euro grant in 2012, together with two other institutes.[21]

The growing interest in systems chemistry stems mostly from the desire to fully understand the origin of life (abiogenesis) – a natural process by which life has arisen from non-living matter.[9,11,15] Biology, with its focus on existing life, has not been able to provide complete answers[9] as such answers require the knowledge of how prebiotic building blocks self-assembled into complex biomolecules, how such molecules were integrated into protocellular entities and how they developed homochirality.[9,11] Such underlying processes likely need to be studied at the level of systems. Systems biology, a sub-discipline focused on molecular interactions within already existing biological networks, has so far been very successful at engineering certain properties *in vivo*,[22] but simultaneously limited in its ability to improve the understanding of the "origins"-related questions.

Systems chemistry presents an alternative approach to the problems faced by its biological counterpart. It studies complex interactions and emergent properties by designing simple chemical networks *de novo* and then progressively increasing their complexity through integration of new components.[9,15] Not only is such a methodology very well applicable to research on prebiotic processes,[23,24] but it also makes the field of systems chemistry conceptually related to synthetic biology as it often results in the invention of alternative biochemistries.[15] In fact, creating subsystems and components for artificial life is another goal of systems chemistry.[9,15] The first chemical model of hemostasis was completed in 2004,[25] a year before the field was formally established. Today, many research groups are working on recreating other important properties of life – self-replication, autoregulation and the ability to function out of equilibrium.[18] As a notable example, chemists from the "Building a Synthetic Cell" consortium (BaSyc) work on non-equilibrium chemical systems that, in the fullness of time, are supposed to comprise artificial metabolism of synthetic cells.[26,27]

3.2. Chemical cycles

Modeling known biochemistries as well as their synthetic alternatives would, in most cases, entail the development and integration of chemical cycles. In terms of chemistry, a cycle is a reaction pathway in which the starting substrate is also the final product. Such a substrate, akin to a catalyst, is not ultimately consumed, allowing the cycle to operate repeatedly. In fact, when Hans Krebs discovered the first metabolic cycle in 1932, known as the ornithine cycle (or urea cycle),[28] he remarked on its capability for producing more than 30 urea molecules at the expense of a single ornithine.[29] As shown in **Figure 1**, there is a number of substrates consumed during the cycle execution (most notably, carbamoyl phosphate and aspartate), but the ornithine always regenerates upon full completion. There is also a number of synthesized molecules that do not undergo any subsequent reactions – their quantity can only increase – and as such, they could be considered as products of the whole cycle.

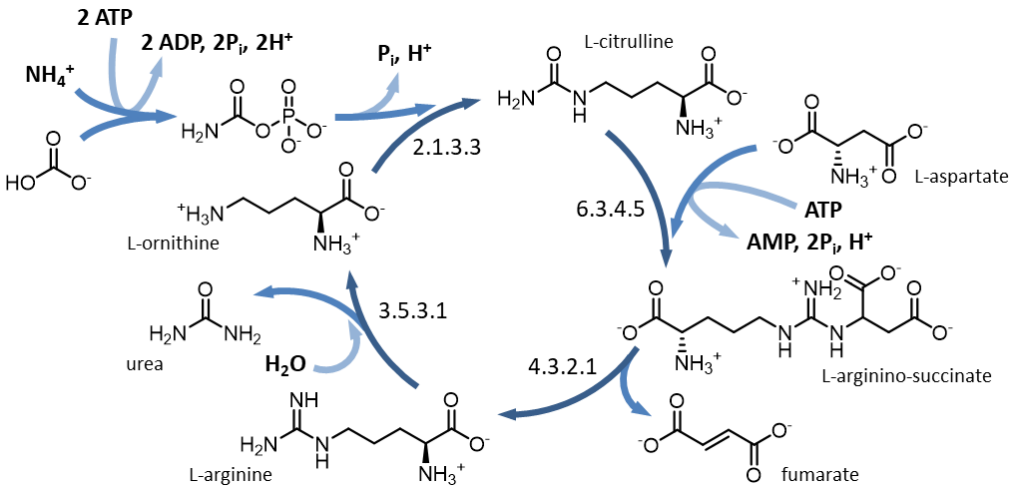


Figure 1. The urea cycle is the first cycle discovered. Its purpose is to transform excess ammonia into its less toxic form – urea. As shown in the left upper corner, the process begins outside of the cycle with creation of carbamoyl phosphate, which then enters the cycle. The numbers shown next to the reaction arrows denote their respective Enzyme Commission numbers.

The next cycle discovered by Hans Krebs, five years later, is of key importance to cell metabolism. Known as the tricarboxylic acid cycle (TCA) or the Krebs cycle (**Figure 2**),[28] it serves as a central hub, interconnecting fat, protein and carbohydrate metabolism. Present in all aerobic organisms, it supplies both energy and building blocks to cell biochemical pathways, which include numerous other cycles.[30,31] For example, while the role of the previously described ornithine cycle is to biosynthesize urea, it also yields various by-products (shown in **Figure 1**) and the Krebs cycle actively participates in recycling these products into substrates, as explained in **Figure 2**. The significant interest in the TCA cycle is well-reflected by the numerous attempts to recreate it *in vitro*, which date back to the 1950s[32,33] and are still being pursued.[34]

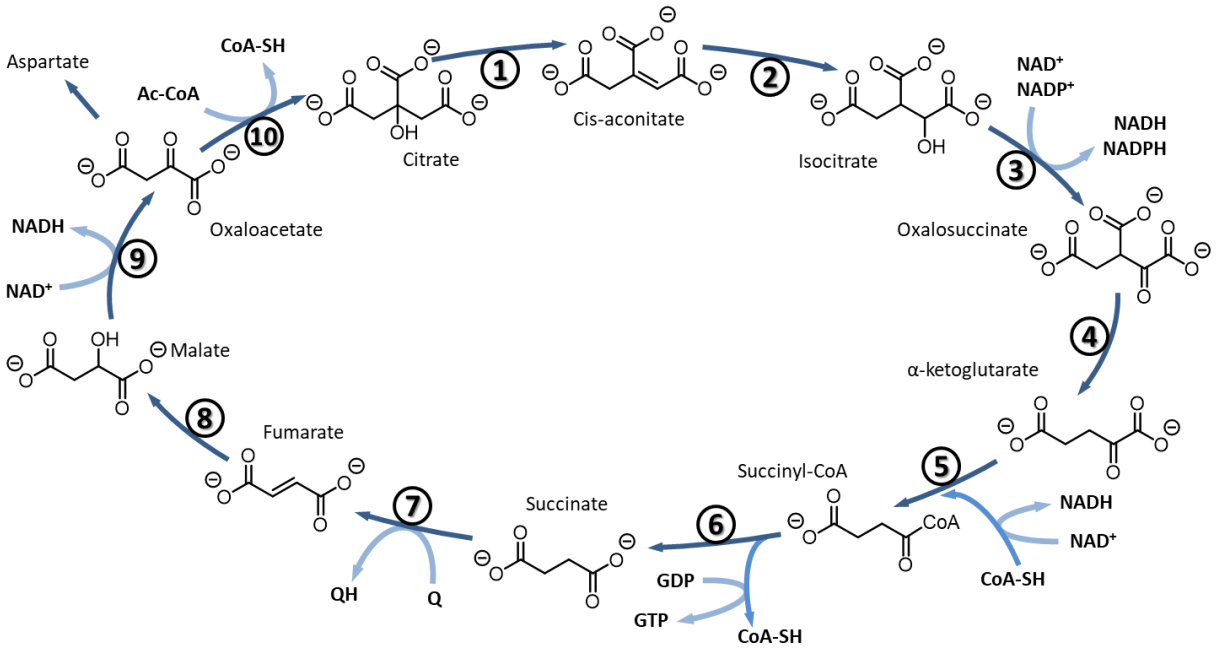


Figure 2. An overview scheme of the Krebs cycle. Steps 8 and 9 are involved in a metabolic pathway that transforms fumarate into aspartate, which is then used in the urea cycle. NADH produced during steps 3, 5 and 9 is used by various metabolic pathways, as well as in the process of phosphorylative oxidation, which transforms ADP back into ATP. CoA = coenzyme A; Q = quinolone.

Some cycles reuse their own by-products. The example shown in **Figure 3** initially uses the adenosyl group from ATP as an activator in spermidine biosynthesis,[35] but instead of later expelling it to be recycled elsewhere, it salvages the carbon atoms from the ribose part to rebuild the L-methionine required for another completion of the cycle.[36]

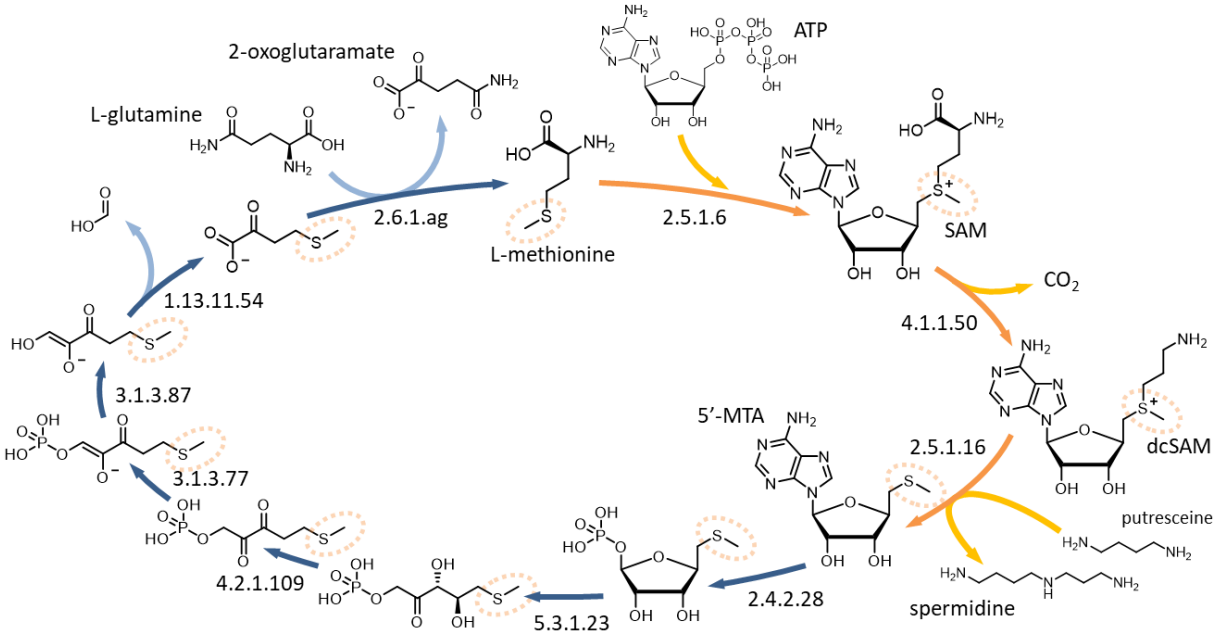


Figure 3. The L-methionine salvage cycle is an intricate system for reusing matter. It synthesizes spermidine, then uses the by-products as precursors of new substrate molecules, allowing itself to synthesize even more spermidine. Adenosine triphosphate, an energy source commonly used in cells, is first used as molecular "template" on which the L-methionine is transformed into a product spermidine in two steps (orange arrows). Next, the S-methyl-5'-thioadenosine, resulting from the same cleavage as the product spermidine, is recycled into L-methionine that reacts with another ATP and the cycle repeats (blue arrows). Thus each molecule of ATP participates (both directly and indirectly) in the synthesis of two molecules of spermidine, staying within the cycle for two completions of the "orange" segment and one of the "blue" segment. The thiomethyl moiety (marked with the orange dotted ovals) is the only group of atoms that always remains in the cycle, whereas all remaining matter passes through the cycle. Inorganic reagents, such as phosphates and hydrogen cations, are omitted for clarity. SAM = S-adenosylmethionine; dcSAM = decarboxylated SAM or S-adenosylmethioninamine; 5'-MTA = S-methyl5'-thioadenosine.

Not every metabolic cycle is supposed to synthesize a target side-product. Some cycles discovered *in vivo* are futile – i.e., their full completion yields no apparent net effect other than dissipation of energy. Despite their name, such "futile" cycles serve many important roles in living organisms. The mannitol cycle presented in **Figure 4** is one such example.[37] It transforms fructofuranose-6-phosphate into D-mannitol, but then transforms it back into the starting furanose, whereby the only apparent effect of the cycle completion is hydrolysis of ATP to ADP, which looks like a waste of precious energy. Mannitol, however, is one of the most abundant storage molecules in nature[38] – completing the first two steps of the cycle results in a formation of a useful energy stockpile in a form that can be transferred between cells.[39]

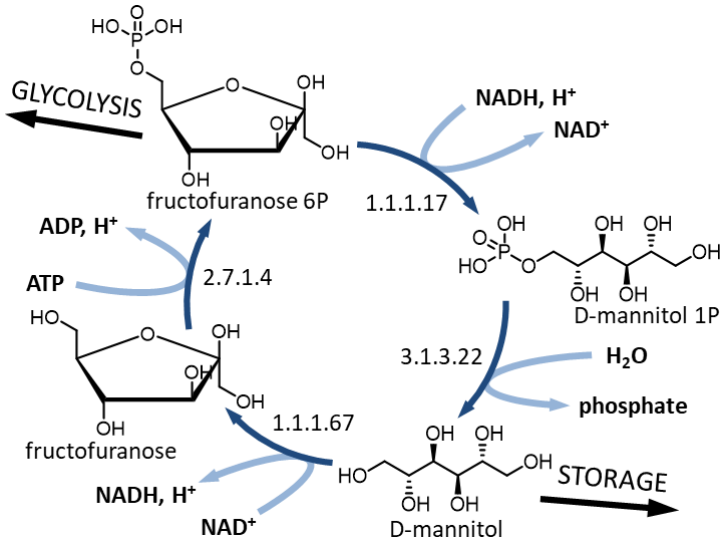


Figure 4. The mannitol cycle is an *in vivo* example of a futile cycle. Even though it does not synthesize any important side-products (hydrolysis of ATP being the only net change), it serves an important role of storing energy in the form of D-mannitol. Fructofuranose-6-phosphate is transformed outside of the cycle into glucose-6-phosphate, which participates in glycolysis. The numbers shown next to the reaction arrows denote their respective Enzyme Commission numbers.

The proper functioning of many metabolic cycles (especially the futile ones) is ensured by partial separation of their components. The urea cycle is split between mitochondria and cytosol,[28] while mannitol in plants is split between leaf cells and sink tissue cells.[39] How far such separation can go is best exemplified in **Figure 5** by the Cori and Cahill cycles, responsible for muscle sores.[40] Depending on the level of muscular fatigue, pyruvate coming from energy-releasing glycolysis is further decomposed into either lactic acid (the Cori cycle) or alanine (the Cahill cycle). These two compounds are then transported with blood all the way from a muscle to the liver, where they are transformed back into pyruvate. This pyruvate undergoes energy-consuming gluconeogenesis and the resulting glucose is sent back to muscles to fuel their intensive work.

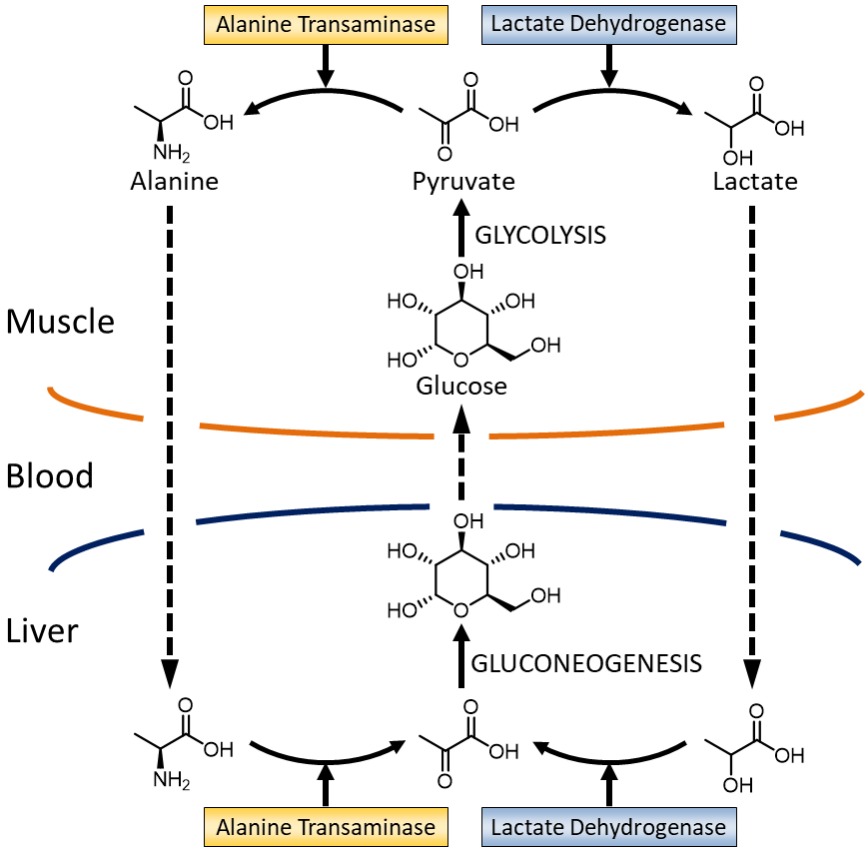


Figure 5. The Cahill and Cori cycles participate in energy transfer between the liver and muscles. The Cahill cycle is shown on the left and the Cori cycle is shown on the right. From a chemical point of view, both cycles are futile. Nevertheless, their components are split into separate phases – glucose consumed in muscles is recovered in the liver, then transported back – thus fulfilling an essential role of supplying working tissues with energy. The solid arrows denote chemical transformations, whereas the dotted arrows indicate the processes associated with transport.

3.3. Chemical replication and auto-amplification

Self-replication is a behavior of non-equilibrium systems capable of constructing their own copies, or copies of some of the system components. In biology, such systems are, for instance, cells reproducing through division[41] or viruses multiplying using infected cell machinery.[42] In biochemistry, it is DNA/RNA strands assembling their copies with the help of polymerases.[28] In systems chemistry, it is one molecule interacting with another to actively promote synthesis of its copies (without being consumed in the process).[43] All self-replicating systems are, by definition, autocatalytic cycles [44] – while the opposite is not always true, a reliable method of assembling new cycles could, by extension, assist in *de novo* design of chemical self-replicators.

Theoretically, if a self-replicating system were allowed to operate continuously (i.e., if it were constantly fed with a sufficient amount of substrates), it should result in an exponential growth in the concentration of the replicated component. In realistic scenarios, however, it is usually formed at a sigmoidal rate as substrates other than the one being replicated run out.[45]

The earliest attempts at creating systems capable of self-replication were heavily based on polynucleotides (building blocks of DNA), reacting in aqueous solutions.[44] The first truly chemical-organic model was created by the Rebek group,[46] who in 1990 synthesized a relatively complex molecule capable of self-replication in non-aqueous solvents. The entire system is shown in **Figure 7** – based on hydrogen bonds just like DNA, it exhibited a sigmoidal time-concentration profile, typical of autocatalytic processes. Such a design of hydrogen-bonding substrates assembling on a molecule identical to the product was successfully applied in multiple follow-up experiments, the two most notable of which are based on the Diels-Alder reaction (**Figure 8a-8c**)[47] and 1,3-dipolar cycloaddition (**Figure 8d-8f**).[48]

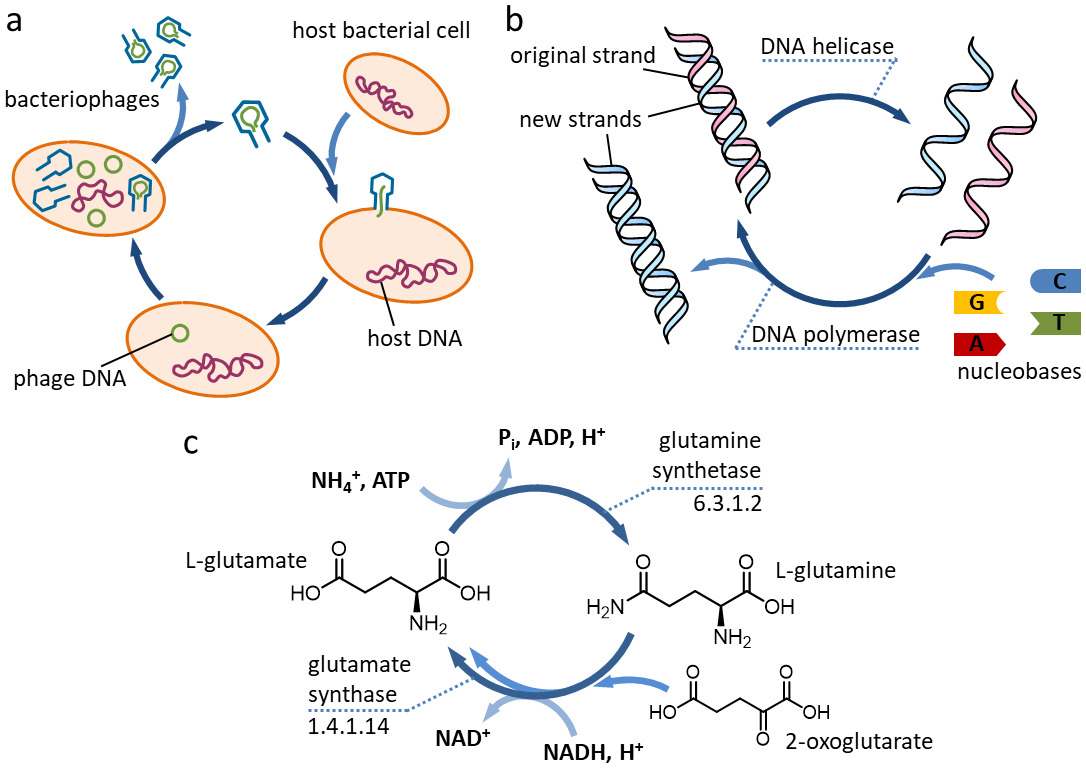


Figure 6. Naturally occurring processes of self-replication. Scheme (a) illustrates viral replication. Starting from the top, a bacteriophage virion infects a bacterial cell with its DNA. The phage DNA then circularizes and replicates, leading to the biosynthesis of the phage proteins by the host. The proteins are assembled into new phages, which are released once the cell lyses. Scheme (b) illustrates a DNA replication process. The original double-stranded DNA is split by helicases and on each of its strands a new strand is being assembled with the use of complex cellular machinery. Both schemes (a) and (b) are intentionally simplified (e.g., polymerases typically build new strands, whilst a DNA double helix is being split). The simple metabolic cycle shown in (c) is known as the ammonia assimilation cycle.[49] In a two-step process, L-glutamate actively promotes synthesis of its copy by transforming into L-glutamine and undergoing a disproportionation reaction with 2-oxoglutarate, resulting in two equivalents of L-glutamate (one copy plus one recovered). The numbers shown next to the reaction arrows denote their respective Enzyme Commission numbers.

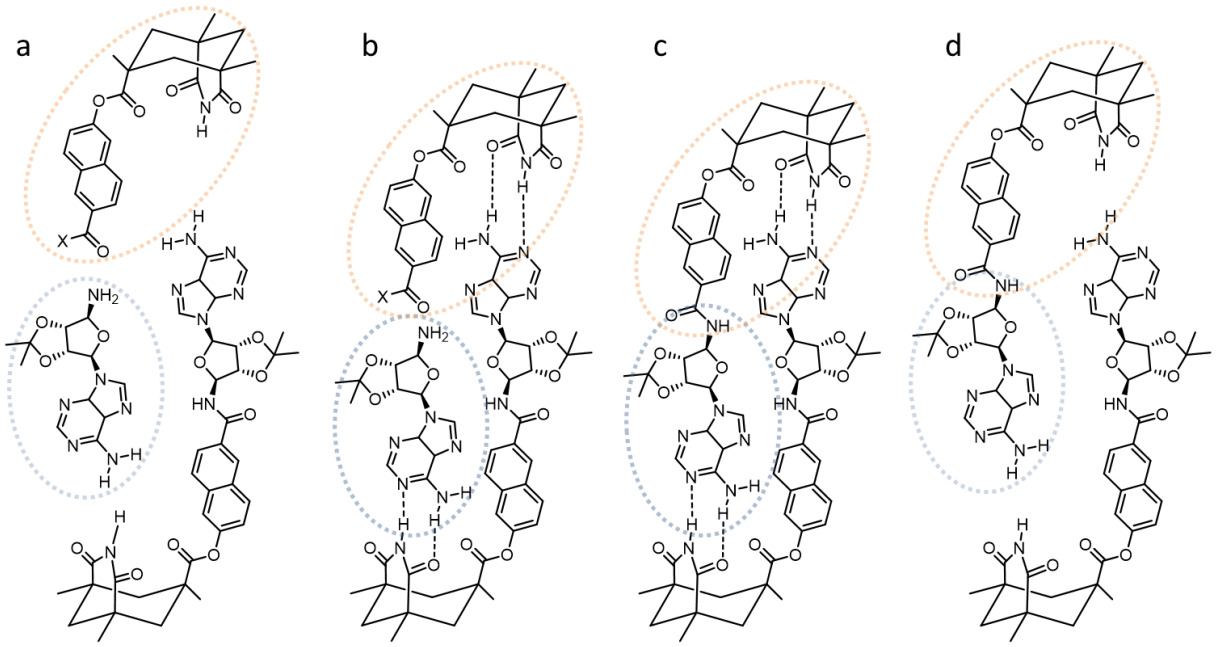


Figure 7. The self-replicating system developed by the Rebek group. Free reactants are seen in (a). The dotted ovals denote substrate molecules, while the catalyst identical with the product of the "incoming" reaction is visible at the bottom right corner of (a). The first step shown in (b) involves hydrogen-bonding of the substrates to the catalyst. This brings them close together, allowing the amide formation reaction to assemble the complex shown in (c), where "X" denotes the pentafluorophenyl leaving group. Once hydrogen bonds are broken as in (d), the molecules are set apart, resulting in two identical copies of the catalyst.

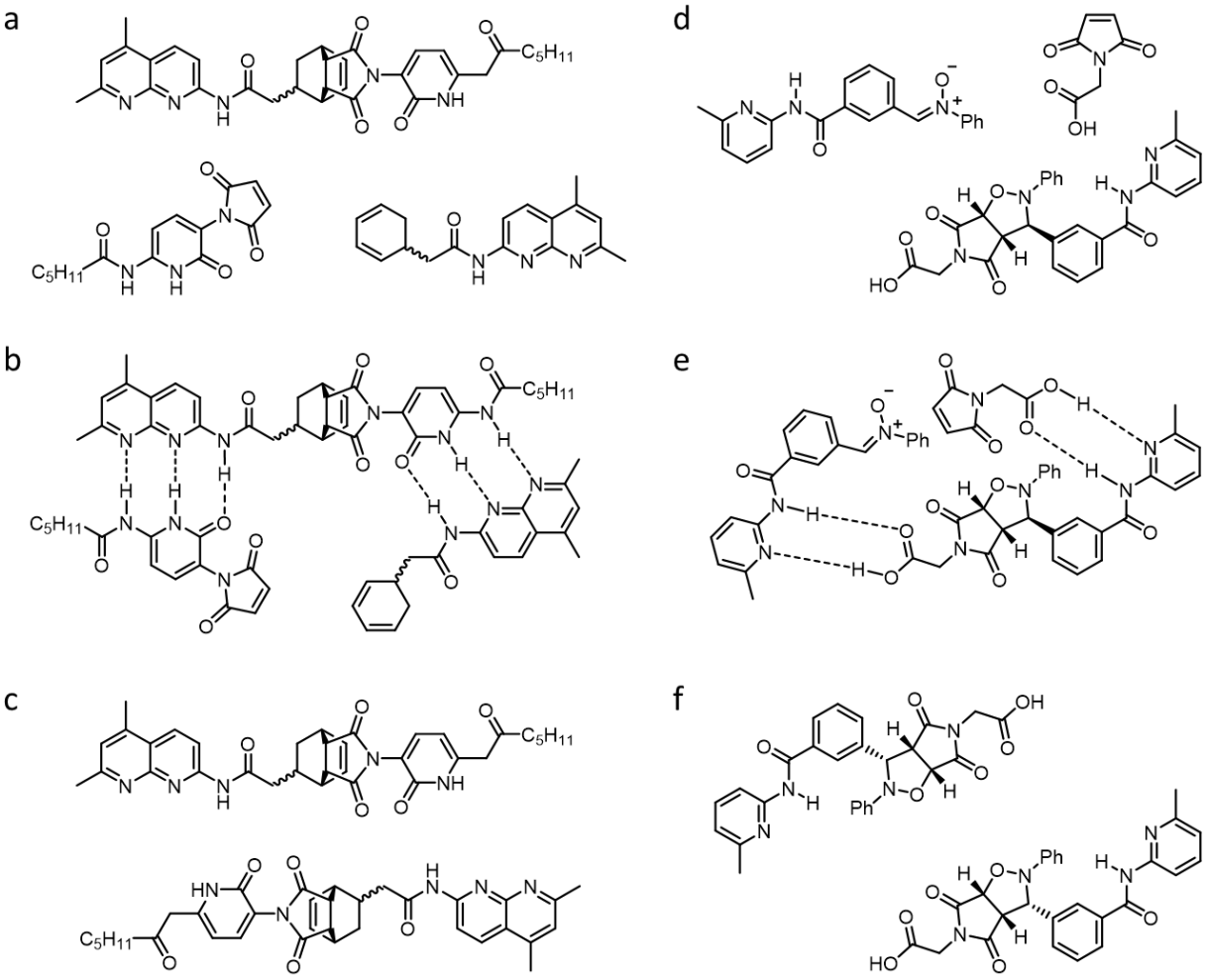


Figure 8. Autocatalytic processes in two self-replicating, chemical systems. The first, shown in schemes (a-c), is based on the Diels-Alder reaction. The substrates shown at the bottom of (a) undergo hydrogen-bonding to a molecular "template", as shown in (b), in a position that promotes the regioselective Diels-Alder reaction, resulting in the copy of the "template" shown in (c). The system shown in schemes (d-f) is based on 1,3-dipolar cycloaddition between a nitrone and maleimide – shown at the top of (d). Hydrogen-bonding of the substrates to the template compound, as illustrated in (e), keeps them sufficiently close and at the right orientation promotes formation of another template molecule, as shown in (f).

Nowadays, continuous flow reactors are used to develop more sophisticated chemical systems, capable of auto-amplification. This is well illustrated by the work of the Whitesides group from 2016, shown in **Figure 9**.[18]

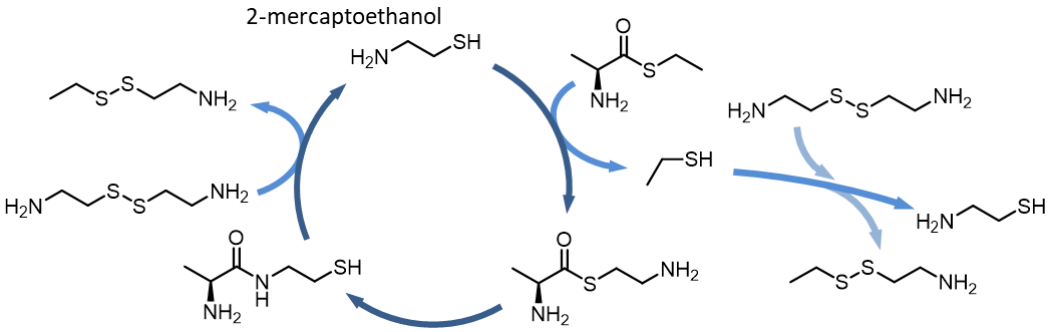


Figure 9. A simplified scheme of a chemical system in which 2-mercaptoetanol is auto-amplified. The system operates in a continuous flow reactor. The reactions responsible for triggers and inhibition are omitted in this scheme.

3.4. Signaling cycle cascades

Just as in metabolism, cellular signaling may also involve reaction cycles. For example, mitogen-activated protein kinase (MAPK) cascades are a diverse family of pathways that pass signals from the cell surface to the nucleus. MAPKs are serine/threonine specific kinases that work in activation loops – catalytically inactive in their base form, they become activated once sufficiently phosphorylated by another kinase, then they deactivate again if dephosphorylated by cytosolic phosphatases. Thus, MAPKs are activated by MAPKKs (MAPK kinases), which are activated by MAPKKKs (MAPKK kinases), which are activated by G-proteins responding to membrane receptors, which, in turn, are activated by extracellular stimulus. Therefore, in the absence of external stimuli all kinases eventually become inactive.

MAPKs are expressed in all eukaryotic cells. Typically, the cascades include three cycles (layers), with sufficient activation (over 1,000-fold) of each kinase being achieved after two phosphorylation events – this architecture is highly conserved from yeast to humans.[50] There are also known atypical MAPK cascades, which comprise only two layers, require a different number of phosphorylations and/or lack some other features[51] – however, in this thesis, we will be concerned only with the "classic" pathways (see **Figure 10**).

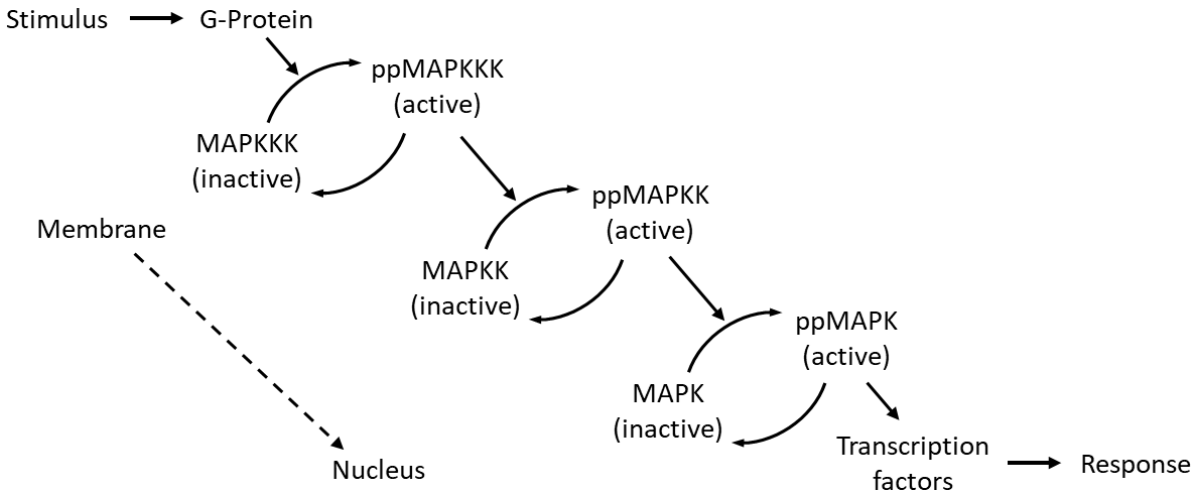


Figure 10. A simplified scheme of a typical MAPK assembly, conserved in most organisms. The dotted arrow denotes spatial orientation. The straight solid arrows denote activation/catalysis. The curved arrows denote activation/deactivation reactions. The state of kinase dual phosphorylation is indicated by "pp".

Mitogen-activated protein kinases are responsible for regulation of multiple processes, including cell division, proliferation, apoptosis and gene expression. Despite their rather specific name, MAPK pathways respond to diverse stimuli – mitogens, growth factors, cytokines, as well as chemical and physical stress.

Every cell employs multiple MAPK pathways – each dedicated to different signals and responses. The assortment of kinases varies strongly between species: plants are known to have a remarkably high number of cascades dedicated to defensive mechanisms (since plants cannot "flee").[50] The signaling cascades present in mammalian cells are presented in **Figure 11** – including the most well-characterized[52] Ras/ERK pathway, which was used as a model in my doctoral research.

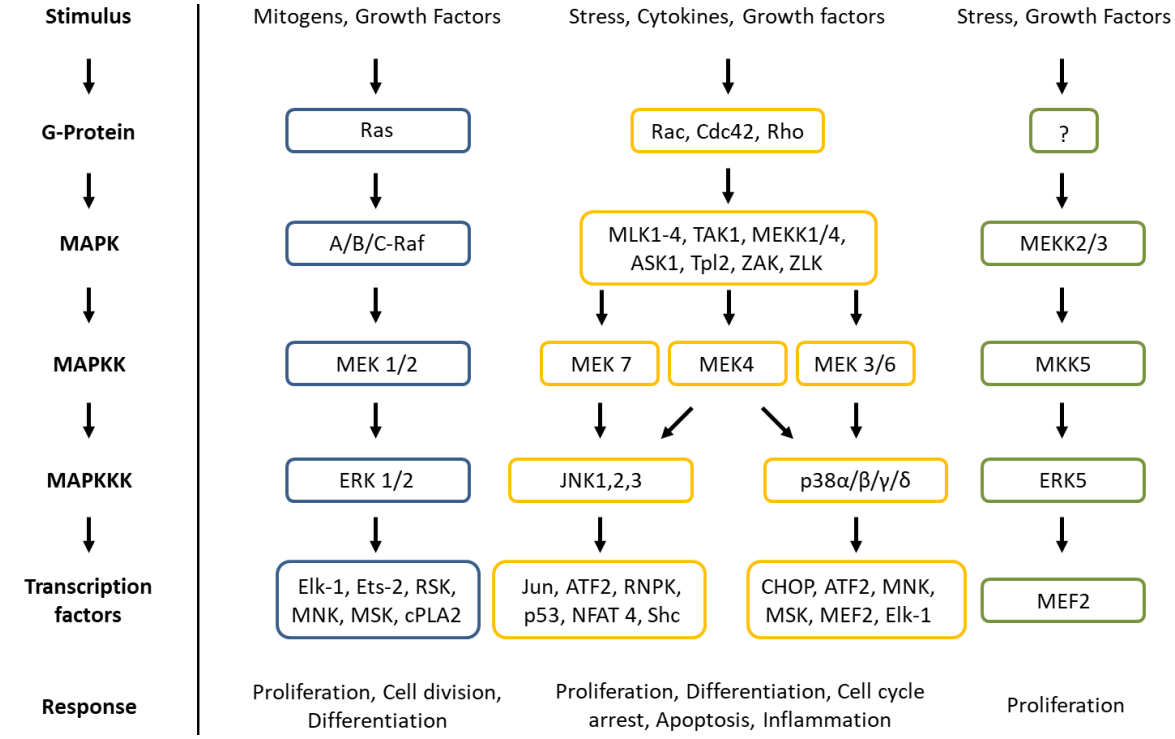


Figure 11. A simplified overview of mammalian MAPK cascades (based on the schemes from Dhillon 2007[53] and Roberts 2007[54]).

As seen from **Figure 11** (orange pathways), some MAPK cascades overlap due to the shared kinases. The specificity of signal transduction is nevertheless maintained with spatial organization[52] and cellular scaffolding.[55] In yeast, MAPK modules are neatly organized by scaffolding proteins (e.g., ste5) that bind multiple kinases simultaneously in order to ensure the specificity of signaling cascades.[56] In neurons, MAPKs are transported along axons by dynein motors in order to make up for long distances.[57] Finally, MAPKs can diffuse through the cytoplasm independently – kinases recruited near the cell membrane activate cytosolic kinases, which travel into the cell interior to activate kinases localized closer to the nucleus, thereby leading to formation of spatial gradients. The strength of such a diffusion-driven signal decreases rapidly with distance unless the cascade has multiple levels – the more levels it has, the farther the signal reaches, which might be the main reason for emergence of "classic" three-layered MAPK cascades.[50,58]

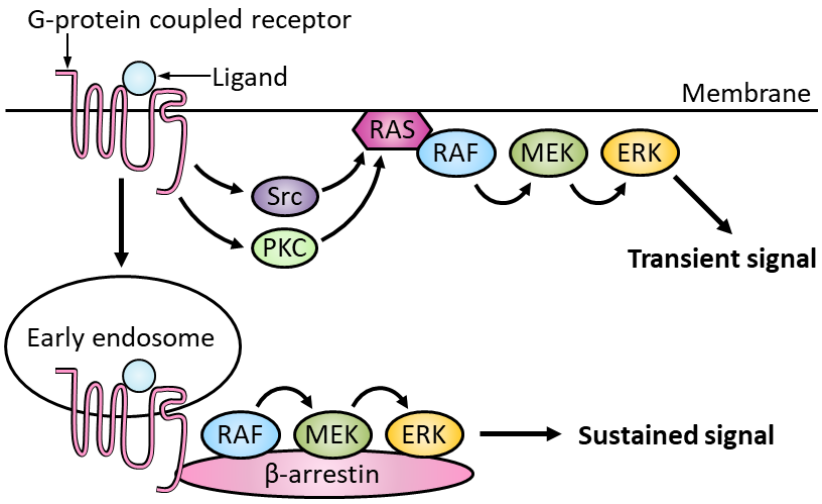


Figure 12. Two modes of the Ras/ERK pathway operation. The top pathway relies on diffusion of kinases and results in a transient signal. The bottom pathway employs the β-arrestin scaffolding protein, resulting in a sustained signal. The scheme presented above was redrawn from Kholodenko (2010).[59]

The timing of MAPK activation is of essence when triggering specific responses. For example, in PC-12 cells, transient activation of ERK induces cell proliferation, while sustained activation induces cell differentiation.[59,60] **Figure 12** shows two temporally separated modes of ERK activation – the sustained mode is reliant on scaffolding protein β-arrestin (in a similar fashion to yeast cascades), while the temporary mode is scaffolding-independent.

Understanding MAPK cascades is of medicinal relevance, particularly in the context of cancer.[53,61] In the Ras/ERK pathway, mutations of Ras are present in 90% of pancreatic cancers, 60% of papillary thyroid cancers and 50% of colon cancers, while mutations of B-Raf are observed in 70% of melanomas and 50% of papillary thyroid cancers. [54]

3.5. Chemical oscillators

The search for alternative biochemistries (specifically the Krebs cycle) has led to the discovery of yet another emergent property of reaction networks – their ability to periodically switch between discrete states, giving rise to oscillations.[11,32,33]

At first, chemical oscillations were observed only in heterogeneous systems and even after the so-called "iodine clock" had been discovered, it was still believed that oscillating reactions in homogeneous systems were impossible and it was Belousov's discovery that finally debunked that belief.[33] Presently known as the Belousov-Zhabotinsky (BZ) reaction, it needed a decade, as well as involvement and popularization by another chemist, Zhabotinsky, to obtain its recognition, but once it did, it was followed by a tide of research on non-equilibrium systems and dynamics.[62]

Chemical oscillators (alternatively labeled as oscillating reactions or in some cases "chemical clocks") are reaction networks that produce intermediates whose concentrations periodically increase and decrease.[43] Unlike the classical pendulum, chemical oscillations do not pass through a system equilibrium point – they are a far-from-equilibrium phenomenon, described by non-equilibrium thermodynamics. Consequently, sustained oscillations require an open system with a continuous flow of reagents.[33]

The oscillatory behavior can emerge from a system as simple as the Brusselator[63] [Brussels+oscillator], which is a purely theoretical model described by the following equations:

$$A \to X, \qquad 2X + Y \to 3X, \qquad B + X \to Y + D, \qquad X \to E,$$

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = 0,$$

$$\frac{d[X]}{dt} = [A] + [X]^2[Y] - [B][X] - [X],$$

$$\frac{d[Y]}{dt} = [B][X] - [X]^2[Y].$$

The BZ reaction itself can be described by the following balance equation:

$$3CH_2(COOH)_2 + 4BrO_3^- \rightarrow 4Br^- + 9CO_2 + 6H_2O,$$

with cerium(III) working as a catalyst. Mechanistically, however, it is a complex interplay of multiple transformations.[64,65] One of the models describing it is known as the Oregonator[66] [Oregon+oscillator]:

$$A + Y \overset{k_I}{\rightarrow} X + P, \qquad X + Y \overset{k_{II}}{\rightarrow} 2P, \qquad A + X \overset{k_{III}}{\longrightarrow} 2X + 2Z,$$

$$2X \overset{k_{IV}}{\longrightarrow} A + P, \qquad B + Z \overset{k_V}{\rightarrow} \frac{1}{2} fY,$$

$$\frac{d[X]}{dt} = k_I[A][Y] - k_{II}[X][Y] + k_{III}[A][X] - 2k_{IV}[X]^2,$$

$$\frac{d[Y]}{dt} = -k_I[A][Y] - k_{II}[X][Y] + \frac{1}{2} f k_V[B][Z],$$

$$\frac{d[Z]}{dt} = 2k_{III}[A][X] - k_V[B][Z],$$

$$\frac{d[A]}{dt} = \frac{d[H^+]}{dt} = 0, \qquad X = HBrO_2, \qquad Y = Br^-, \qquad Z = Ce^{IV},$$

$$A = BrO_3^-, \qquad P = HOBr.$$

An example of a contemporary chemical oscillator has already been shown in **Figure 4**. Another system, developed by the Huck group for the purpose of artificial cell research, is shown in **Figure 13**.[17]

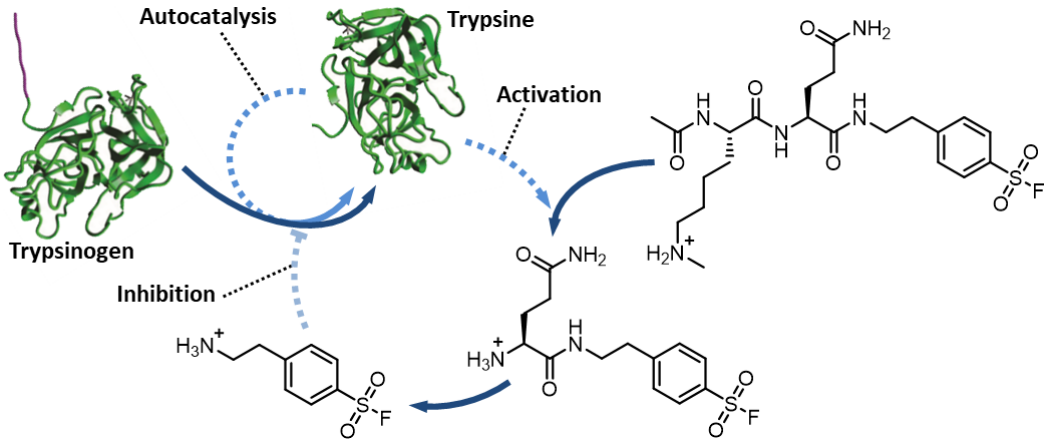


Figure 13. The Huck oscillator. The dark-blue solid arrows denote reactions, whereas the light-blue dotted arrows denote catalysis/inhibition. In this cycle, trypsin promotes both its own synthesis from trypsinogen (auto-amplification) and the synthesis of an inhibitor that negatively regulates trypsin synthesis (auto-regulation), resulting in periodic increase and decrease in trypsin concentration at the outflow detector.

When an oscillating reaction is left unstirred, not only does it demonstrate the behavior described above, but it also gives rise to spatial inhomogeneities, which can be observed as travelling fronts (e.g., chemical waves or Turing patterns).[43] This behavior is analogous to calcium oscillations/waves – ubiquitous signals present in all cells – responsible for transmission of intracellular information. These signals range in frequencies from 10^{-5} to 100 Hz[67] and regulate the processes of cell proliferation, differentiation and death as well as higher functions like fertilization, muscle contraction, learning and memory (see **Figure 14**).[68,69] The activity of Ras protein from the mammalian Ras/ERK cascade is strongly modulated by oscillations of cellular Ca^{2+}, hence many calcium regulatory properties.[70]

Typically, in resting cells, the concentration of intercellular free calcium is maintained at 100 nM. Upon stimulation, this concentration increases up to 1 µM from both the internal and external sources.[28,69] Entry from outside the cell is enabled by the membrane Ca^{2+} channels, controlled by ligand binding (including the drug NMDA), stretch or depolarization.[71] Internal stores include the endoplasmic reticulum (the sarcoplasmic reticulum in muscles), from which Ca^{2+} are released through various channels, most notably 1,4,5-triphosphate and ryanodine receptors.[71] The formation of waves occurs due to the pulsatile manner in which calcium is being released – as concentration increases locally, spatial inhomogeneity, travelling by means

of diffusion, is formed. Interference of multiple, both internal and external, Ca^{2+} waves results in a signal exhibiting a complex spatial pattern, which is recognized and interpreted (decoded) by the signaling machinery of the cells in order to yield a specific response.[42,70] The past research has shown that the activity of calcium-dependent Calmodulin-kinase II – just like release of transmitters into a neuron synapse – is modulated by the frequency of a signal rather than its amplitude or average strength.[72]

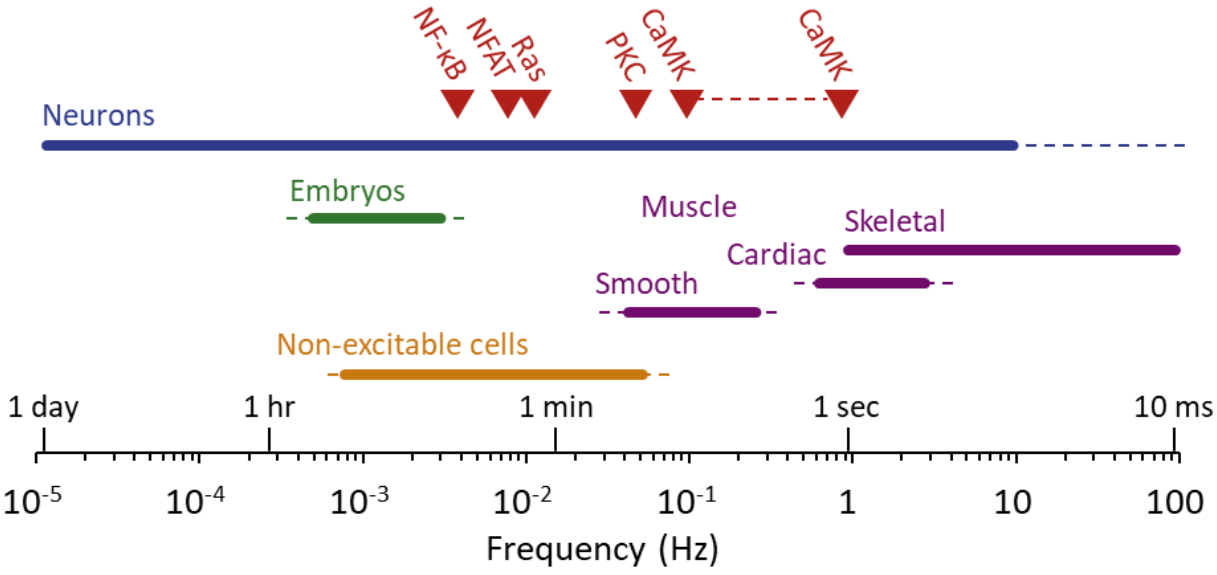


Figure 14. Frequency decoding of physiological Ca^{2+} signals. The red arrowheads denote optimal frequencies for half-maximal activation of specific proteins. The scheme is redrawn from Marchant (2008).[67]

The mechanism by which an oscillating signal is interpreted (decoded) by biochemical machinery is still poorly understood. The Wang group noticed in their theoretical research[73] that there is a strong discrepancy in how maximum and average Ras activity depends on Ca^{2+} frequency, but they did not investigate it any further.

Figure 14 illustrates that the Ras protein of the Ras/ERK cascade is strongly activated by a Ca^{2+} signal of 0.01 Hz. However, this information refers to certain published experimental data, which requires some explanation.[74] During the experiment, the Ca^{2+} frequency gradually changed from 0.00167 Hz to 0.0167 Hz. A significant Ras activation was observed at 0.0083 Hz and kept increasing until the boundary of 0.0167 Hz was reached. Thus, the exact maximal activation frequency remains uncertain.

3.6. Enzymatic reactions

3.6.1. Michaelis-Menten kinetics

The study of system oscillatory and/or autocatalytic behavior, as well as some other potential emergent properties, entails the analysis concerning reaction kinetics of the system. While chemical networks typically obey n^{th} order kinetic equations which, as seen from Brusselator, are not necessarily complex, biochemical networks (e.g., the metabolic cycles shown in section **3.2** and the cascades shown in section **3.4**) depend upon enzymes to catalyze their own reactions and thus follow either the Michaelis-Menten or Hill rate equations. Enzymatic kinetics are different from chemical in that they are always non-linear and characterized by a maximum reaction rate – it is not possible to exceed this limit by increasing the substrate concentration. They also introduce a new parameter, the Michaelis constant, to describe how enzymatic reaction rates are significantly limited by the substrate-binding step.

The models of kinase cascades used in my doctoral research are described later, in sections **5.3** and **5.5,** by the Michaelis-Menten rate equations – before those cascades could be analyzed at the network level, just like their chemical counterparts, their non-linear reaction kinetics had to be addressed first (sub-section **5.4.1**).

According to Beard (2008)[75] the simplest model case of Michaelis-Menten kinetics could be described by the following equations:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} E + P,$$

$$\frac{d[E]}{dt} = (k_{-1} + k_2)[ES] - k_1[S][E], \qquad \frac{d[ES]}{dt} = k_1[S][E] - (k_{-1} + k_2)[ES],$$

where *E* is the free enzyme, *S* is the substrate, *ES* is the enzyme-substrate complex and *P* is the product. The total concentration of enzymes $E_0 = [E] + [ES]$ is constant.

The rate of product formation is found by using a steady-state approximation:

$$\frac{d[E]}{dt} = \frac{d[ES]}{dt} = 0,$$

$$[E] = \frac{k_{-1} + k_2}{k_{-1} + k_2 + k_1[S]} E_0, \qquad [ES] = \frac{k_1[S]}{k_{-1} + k_2 + k_1[S]} E_0,$$

$$\frac{d[P]}{dt} = k_2[ES] = k_2 \frac{k_1 E_0[S]}{k_{-1} + k_2 + k_1[S]} = \frac{k_2 E_0[S]}{(k_{-1} + k_2)/k_1 + [S]}.$$

The last expression can be further simplified by introducing two parameters: maximum velocity $V_{max} = k_2 E_0$ (where k_2 is also known as the turnover number – a number of conversions done by one molecule of an enzyme in one second) and the Michaelis constant $K_M = \frac{k_{-1}+k_2}{k_1}$, which equals concentration of *S* at which the reaction proceeds at half of its maximum velocity. The result is known as the Michaelis-Menten equation:

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{K_M + [S]}.$$

While V_{max} relates to both the amount and catalytic power of an enzyme, K_M relates to the binding-dissociation rate of the reactant. Multiple Michaelis constants may appear in an expression – one for each substrate that may occupy the active site. This is best illustrated by the model of reversible enzymatic catalysis, where two opposite reactions compete for the enzyme active site:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} E + P,$$

$$\frac{d[E]}{dt} = (k_{-1} + k_2)[ES] - k_1[S][E] - k_{-2}[P][E],$$

$$\frac{d[ES]}{dt} = k_1[S][E] - (k_{-1} + k_2)[ES] + k_{-2}[P][E].$$

This is a somewhat more accurate model of enzymatic catalysis, as the product can also bind to the active site and a reverse transformation may occur.

Again, for $E_0 = [E] + [ES] = const$ and $\frac{d[E]}{dt} = \frac{d[ES]}{dt} = 0$:

$$[E] = \frac{k_{-1} + k_2}{k_{-1} + k_2 + k_1[S] + k_{-2}[P]} E_0, \qquad [ES] = \frac{k_1[S] + k_{-2}[P]}{k_{-1} + k_2 + k_1[S] + k_{-2}[P]} E_0,$$

$$\frac{d[P]}{dt} = \frac{(k_1 k_2[S] - k_{-1}k_{-2}[P])E_0}{k_{-1} + k_2 + k_1[S] + k_{-2}[P]} = \frac{k_2 E_0 \frac{k_1}{k_{-1} + k_2}[S] - k_{-1}E_0 \frac{k_{-2}}{k_{-1} + k_2}[P]}{1 + \frac{k_1}{k_{-1} + k_2}[S] + \frac{k_{-2}}{k_{-1} + k_2}[P]}.$$

This time, four parameters are introduced – $V_{max}^{fw} = k_2 E_0$, $V_{max}^{rev} = k_{-1} E_0$, $K_{M,S} = \frac{k_{-1}+k_2}{k_1}$ and $K_{M,P} = \frac{k_{-1}+k_2}{k_{-2}}$:

$$\frac{d[P]}{dt} = \frac{\frac{V_{max}^{fw}[S]}{K_{M,S}} - \frac{V_{max}^{rev}[P]}{K_{M,P}}}{1 + \frac{[S]}{K_{M,S}} + \frac{[P]}{K_{M,P}}}.$$

The experimental data from enzyme kinetic assays is mostly described with turnover numbers, the Michaelis constants and/or equilibrium constants. When working with Michaelis-Menten kinetics and using experimental parameters from scientific literature and databases, the Haldane relationship can be conveniently used to deduce forward and/or reverse rates:

$$\frac{V_{max}^{fw}/K_{M,S}}{V_{max}^{rev}/K_{M,P}} = \frac{k_1 k_2}{k_{-1}k_{-2}} = K_{eq}.$$

3.6.2 Hill kinetics

The top layer of cascade in section **5.5**, i.e., a cycle that decodes Ca^{2+} waves, features Hill kinetics. The Hill rate equations are used to describe the situations, when enzymes bind multiple substrate molecules at once and each consecutive molecule further facilitates the binding of another. This phenomenon is called enzyme cooperativity and can be explained by a simple model of an enzyme that has two binding sites for substrate *S*:

$$E + S \underset{\longleftarrow}{\overset{K_1}{\longrightarrow}} ES_1, \qquad ES_1 + S \underset{\longleftarrow}{\overset{K_2}{\longrightarrow}} ES_2,$$

$$K_1 = \frac{[ES_1]}{[E][S]}, \qquad K_2 = \frac{[ES_2]}{[ES_1][S]}, \qquad E_0 = [E] + [ES_1] + [ES_2],$$

where K_1 and K_2 are equilibrium constants. Based on these constants, the fraction f_b of unoccupied binding sites can be described as a function of [*S*] with the following expression:

$$f_b = \frac{[ES_1] + 2[ES_2]}{2E_0} = \frac{K_1[S] + 2K_1K_2[S]^2}{2(1 + K_1[S] + K_1K_2[S]^2)},$$

which is known as the Hill equation. In one special case, when both sites are identical and independent of each other:

$$K_1 = 2K,$$

because there are two sites available for binding the first substrate and

$$K_2 = K/2,$$

because both sites can release the substrates in a reverse process. The expression for f_b is simplified to the form:

$$f_b = \frac{K[S]}{1 + K[S]}.$$

Since the reaction rate is linearly proportional to the fraction of occupied sites:

$$\frac{d[P]}{dt} = V_{max} f_b = \frac{V_{max}[S]}{1/K + [S]},$$

its expression becomes the Michaelis-Menten equation, described in the previous section.

In another special case, the two binding sites are fully cooperative – either both bind the substrate or none does:

$$K_1 \approx 0, \qquad K_2 \approx \infty, \qquad K_1 K_2 = K^2.$$

The resulting expression for f_b is as follows:

$$f_b = \frac{K^2[S]^2}{1 + K^2[S]^2}.$$

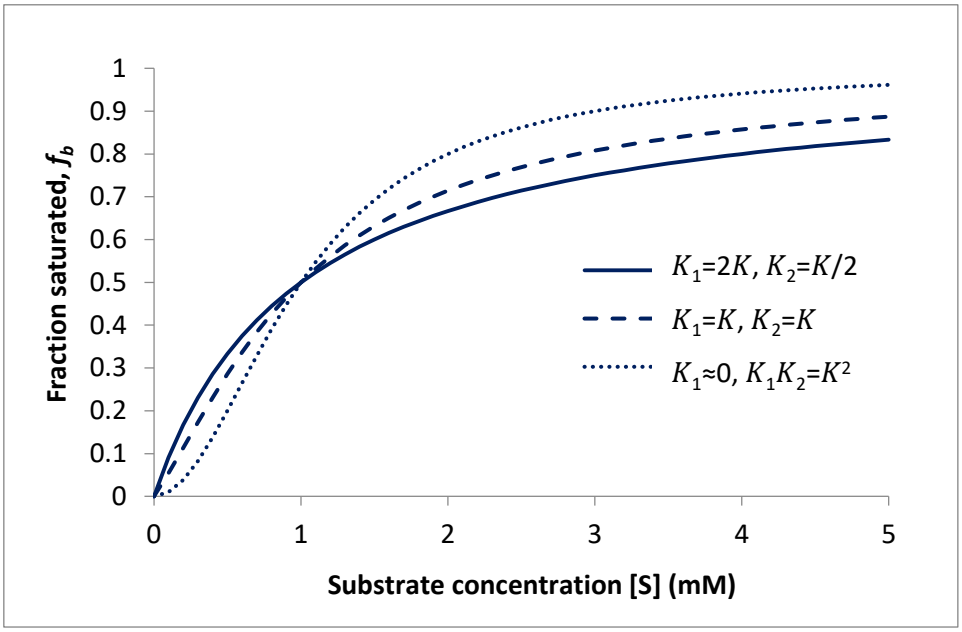


Figure 15. The fraction of saturation f_b as a function of substrate concentration [S] for an enzyme with two substrate-binding sites. The plots are for $K = 1\text{mM}$. The solid plot represents two independent sites and the dotted plot represents two fully cooperative sites.

The shapes of both $f_b([S])$ functions are plotted in **Figure 15**. It can be seen from there that all curves cross at the same point of 50% saturation. The level of enzyme cooperativity can be quantified in terms of steepness at $f_b = 0.5$. The parameter used to characterize this steepness is known as the Hill coefficient, defined as:

$$h = 2\left(\frac{d \ln f_b}{d \ln[S]}\right)_{f_b=0.5}.$$

If $h > 1$, cooperativity is positive. If $h < 1$, cooperativity is negative. If $h = 1$, all sites are independent.

In the first special case – $K_1 = 2K, K_2 = K/2 - h = 1$, corresponding to no cooperativity. In the second special case – $K_1 \approx 0$, $K_2 \approx \infty -$ $h = 2$, corresponding to the theoretical maximum for the case of two binding sites. However, in cases when there are more binding sites, the Hill coefficient *h* might have a higher value. The general Hill equation is as follows:

$$f_b = \frac{K^h[S]^h}{1 + K^h[S]^h}.$$

Since $f_b = \frac{v}{V_{max}}$, the general rate equation for an enzymatic one-substrate one-product reaction is as follows:

$$\frac{d[P]}{dt} = V_{max} f_b = \frac{V_{max} K^h[S]^h}{1 + K^h[S]^h} = \frac{V_{max}[S]^h}{K_M + [S]^h}.$$

4. Cyclorg-based research

4.1. Introduction

Cyclorg software, which is one of two main topics in this thesis, can be construed as the initial step towards systematic discovery and design of chemical cycles. To be clear, it is not a *de novo* prediction tool and does not have mechanistic understanding of the chemistry involved – it is, however, based on the validated literature precedents and can efficiently navigate the output of the published organic reactions (the so-called Network of Organic Chemistry, NOC[1]), assembling them into cycles, as illustrated in **Figure 16**. It does so according to the structures of the chemicals involved – two transformations can be incorporated into a cycle if structures of the product of the first reaction and the substrate of the second reaction are both identical. All (even slightly) imperfect matches are omitted. This is certainly a limitation because with all the variety of organic molecules – homology, constitutional and configurational isomerism – one could conceive many more analogous cycles made up from similar molecules and/or reactions. By limiting ourselves only to exact literature precedents and exact reaction matches, the chances of exactly the same molecule being used in two independent reactions are greatly reduced and differences that would be negligible in the eyes of a trained chemist – a bromine atom instead of chlorine or an additional methyl group far from the reactive site – are not accepted by the algorithm. On the other hand, the cycles that we identify are guaranteed to be chemically feasible as, again, every single reaction within them has been validated by experiment. Of course, because Cyclorg is based on literature precedents, it is also sensitive to errors with which such reactions were catalogued in the reaction/patent databases from which the NOC was derived[76,77] – consequently, one should expect that some of Cyclorg predictions will be false positives. Finally, the software does not take into account reaction kinetics, since such data is largely unavailable.

Notwithstanding such significant limitations, the hypothesis we set out to test using Cyclorg was that "**independent chemists working at different times have jointly – yet often unknowingly – created examples of cyclic reaction sequences**." In order to validate this hypothesis, my task was to identify the most chemically appealing examples from amongst millions of algorithmically generated results.

While the algorithms seeking any and all cycles within the NOC were written by Dr. Piotr Dittwald, my specific focus was on detecting cycles characterized by some purpose, a potential application or another property of interest. In this context, it is significant that most of Cyclorg results were not of such interest – the collection contained many futile cycles of esterification/hydrolysis, dimerization/cleavage or protection/deprotection – "discovering" such examples proves little, even though futility is often seen in living organisms and does not exclude usefulness. The Mannitol (**Figure 7**), Cori and Cahill (**Figure 8**) cycles are both futile, yet they participate in energy management run by biological systems. In the realm of organic chemistry, however, such systems are rather trivial. Nevertheless, they are left available in the final version of Cyclorg as some of them might actually have potential for emergent properties.

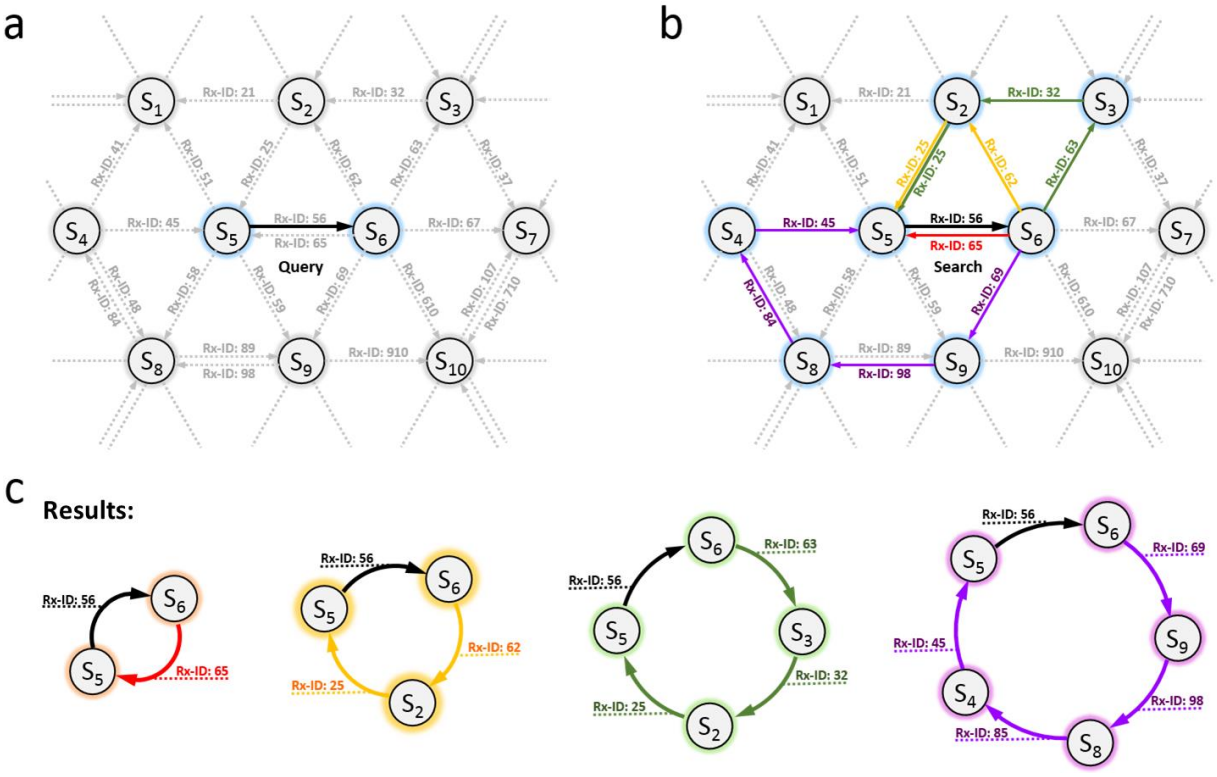


Figure 16. The scheme illustrating the algorithm used to identify reaction cycles within the NOC. a) The algorithm selects reaction $S_5 \rightarrow S_6$ (black arrow) linking substrate S_5 with product S_6. b) Each "backward" path from S_6 to S_5 of desired length (here, L–1 = 1, 2, 3, 4) is identified using standard depth-limited searches. c) Together with initial reaction $S_5 \rightarrow S_6$ these paths create cycles of lengths L= 2, 3, 4, 5. The algorithm is repeated for each possible pair S_i and S_j in the NOC. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.[19]

Still, the software identified a very large number of cycle candidates and the question arose how to inspect them. Since the analysis of the algorithm raw output in the form of SMILES strings (**Figure 17a**) was extremely tedious, I proposed an alternative in the form of graphical representation (**Figure 17c**), which has subsequently been implemented in the final version of Cyclorg.

This representation allows for much faster inspection and also identification of invalid or chemically nonsensical cycles whereby some of the reactions from the underlying literature records miss the key substrates or products. Such problems may reflect manual entry errors of the underlying reaction records during digitization of publications or patents, or they might reflect the functioning of the cycle-search algorithm we have developed. In the case of former class, one example is when the main substrate was omitted and, apparently and quite nonsensically, the reaction record prescribes, say, water producing a complex molecule. In the case of the latter class, a common situation is mistaking counter-ions from reagent salts for actual substrates and products – even if they do not undergo any change during the reaction. Such an example is illustrated in **Figure 18**.

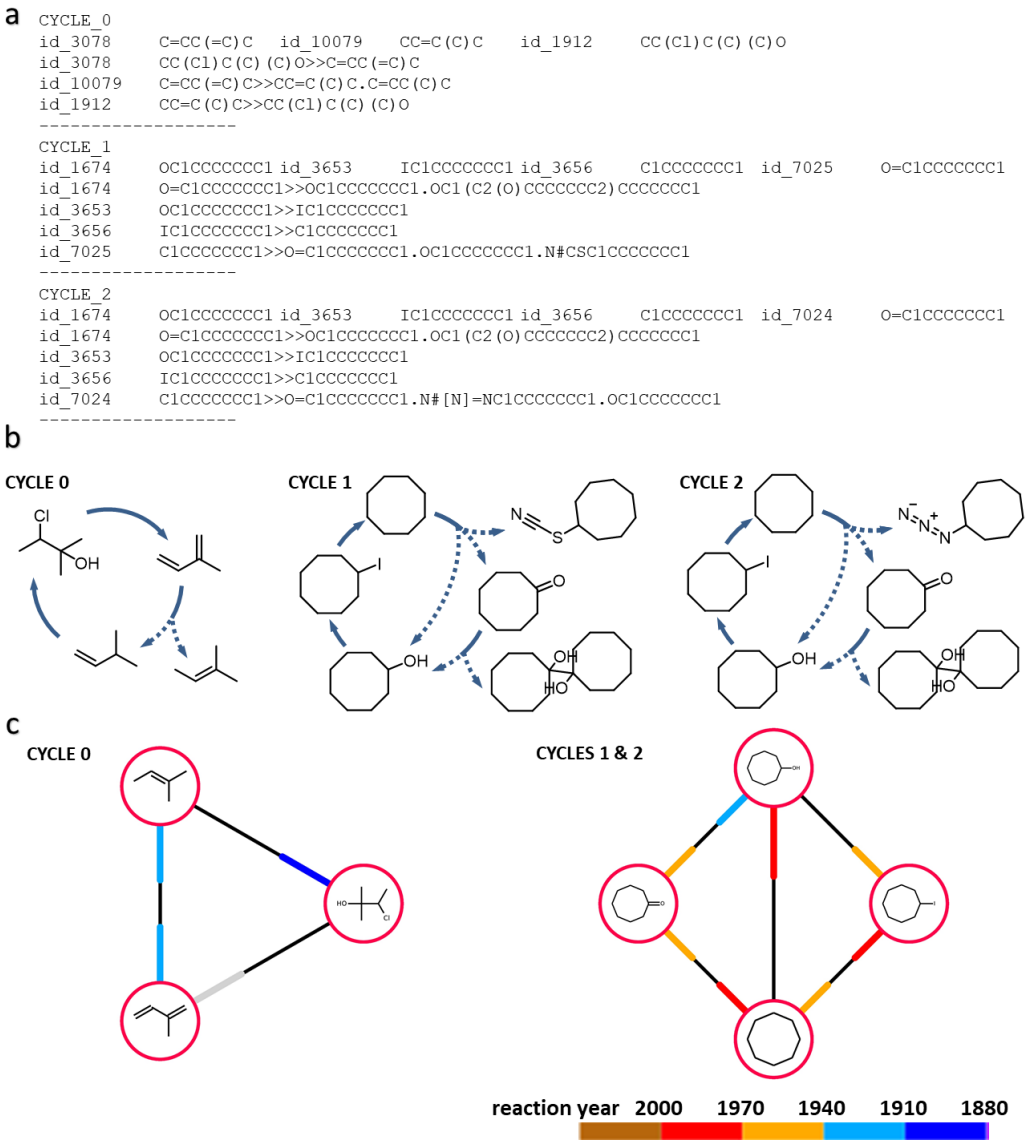


Figure 17. (a) Three examples of cycle candidate output returned by the algorithm in text representation, as SMILES strings. These cycles are redrawn in (b) with the dotted arrows denoting reactions that result in mixtures of the products. (c) Graphical representation of the cycles used in Cyclorg. The colored endings of the connections denote the reaction directions (some are bidirectional) and the year of the first publication (see the legend at the bottom). Cycles 1 and 2 appear on the same graph since side products are not shown in this representation. Such a solution allows for including all literature entries, conditions and reactant combinations in a single comprehensive entry. Nevertheless, “internal connections” are still shown (examples 1 and 2). The cycles shown in (a) and (c) belong to different (now obsolete) collections, thus their id numbers may vary from those stored by Cyclorg.

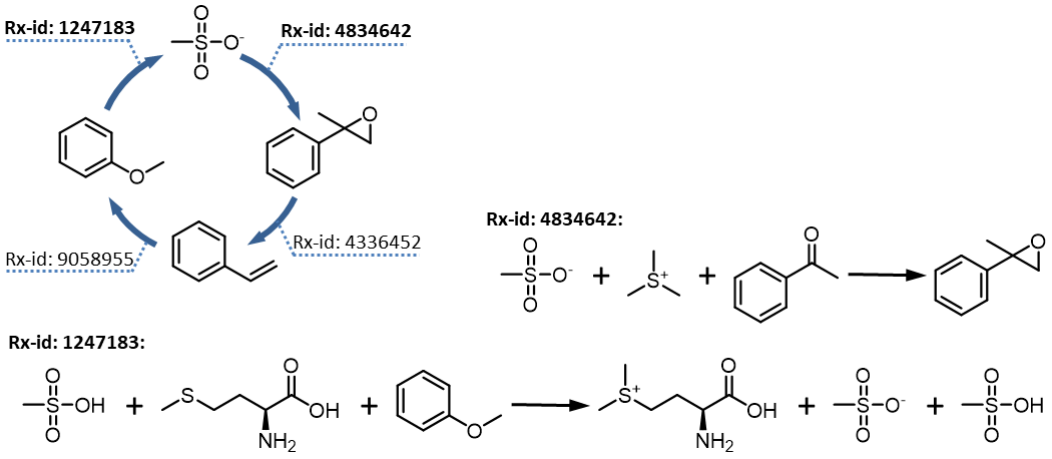


Figure 18. The example mistakenly identified as a cycle is actually a linear pathway starting from acetophenone and leading to S-methyl-L-methionine methanesulfonate. The reaction equations are shown as they were entered in the NOC. The methanesulfonate anion, which appears in the top vertex, is present only to form salts – apart from the acidic proton, it does not exchange any atoms with other reagents. It is, however, present in the NOC's reaction equations. Since the algorithm does not keep track of atom migration or atom balance, it accepted the counter-ion as an actual substrate/product without recognizing its simultaneous presence on both sides of the reaction arrow.

In order to solve such problems, we initially modified the search algorithm to allow only the largest substrate and the largest product of each reaction to become intermediates in the cycle (i.e., to appear in the vertices of a cycle), as opposed to becoming side-products and substrates. With this narrowing criterion, the searches still produced over 6 million cycles, but the vast majority of them were now chemically valid. However, most of the cycles did not seem to exhibit useful functions, were futile and/or only featured interconversions of functional groups. It turned out that the narrowing criterion was too stringent as all the cycles that produce large compounds from smaller parts got excluded during the search – and these are arguably the most valuable from the point of view of an organic chemist.

Accordingly, a new solution had to be implemented to enable synthetically useful cycles. The previous restriction on smaller reactants was lifted – this time all substrates and products were allowed, but in order to prevent reappearance of invalid candidates, the lists of common solvents and inorganic reagents (shown in **Figure 19**) were manually curated, and the algorithm then used these lists to filter such spurious substances off. In this way, over

18 million algorithmically-generated cycles were retained in the new database. It is from this collection that cycles ultimately published in our *Angewandte Chemie* paper[19] were selected. To facilitate the searches, I was also able to define a new set of filters that prevent many useless and uninteresting candidates from appearing in the search results of Cyclorg.

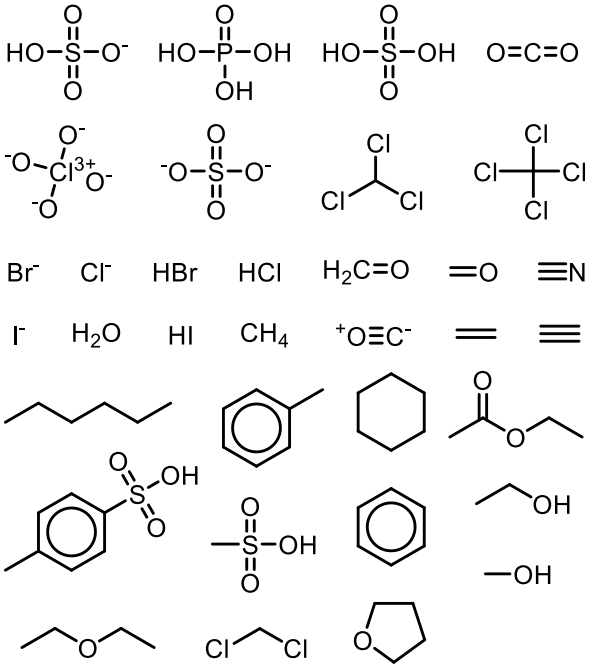


Figure 19. Common solvents and reagents that have been excluded by the algorithm during generation of “side substrate/products allowed” databases.

4.2. Details of the cycle identification

When Cyclorg displays the cycles as graphs such as those shown in **Figure 17c**, only their intermediates are detailed (i.e., the molecules involved directly in the looped pathway, or the cycle "core"– the ones that regenerate after each completion). The reaction conditions and remaining reagents have to be identified from literature. At the time, rx-id numbers were provided for every transformation, allowing for a rapid access to relevant entry in Reaxys (see **Figure 9b**).[78] Such implementation provides a necessary clarity, allowing for a larger quantity of the cycles to be displayed in Cyclorg UI screen.

The major role of Cyclorg is to serve as an "idea generator" for the cycles – in doing so, the software does not admit major changes in the structures of the cycle intermediates, but allows for significant flexibility in altering the structure of the side-products beyond the specific molecules present in the original NOC collection.

The process of cycle identification is illustrated in **Figure 20**. Once the search is finished, the returned cycles are displayed on the screen as graphs arranged in a grid (example of such a graph was presented in **Figure 17c**). At this stage, the potentially interesting cycles (according to the user's preference) are recognized only by the structure of the intermediate molecules and the order in which they are transformed (as shown in **20a**). Such preliminary candidates are then inspected individually, according to the source chemical literature – the remaining reagents are determined at this stage (as presented in **20b**). At the time, Cyclorg provided rx-ids for all the transformations involved in the inspected cycles – and by extension, proposed specific reactants and conditions. In case of some cycle candidates, additional literature studies allowed for discovering alternative conditions, superseding the ones originally proposed by the software (as exemplified in **20c**).

Querying for the substrate and the product of a reaction involved in the cycle within Reaxys[82] gives an instant access to a range of available (and verified) reaction by-products, building blocks and reaction conditions – allowing for redesigning of the cycle and/or "updating" it to more contemporary chemistry (as exemplified in **20c**).

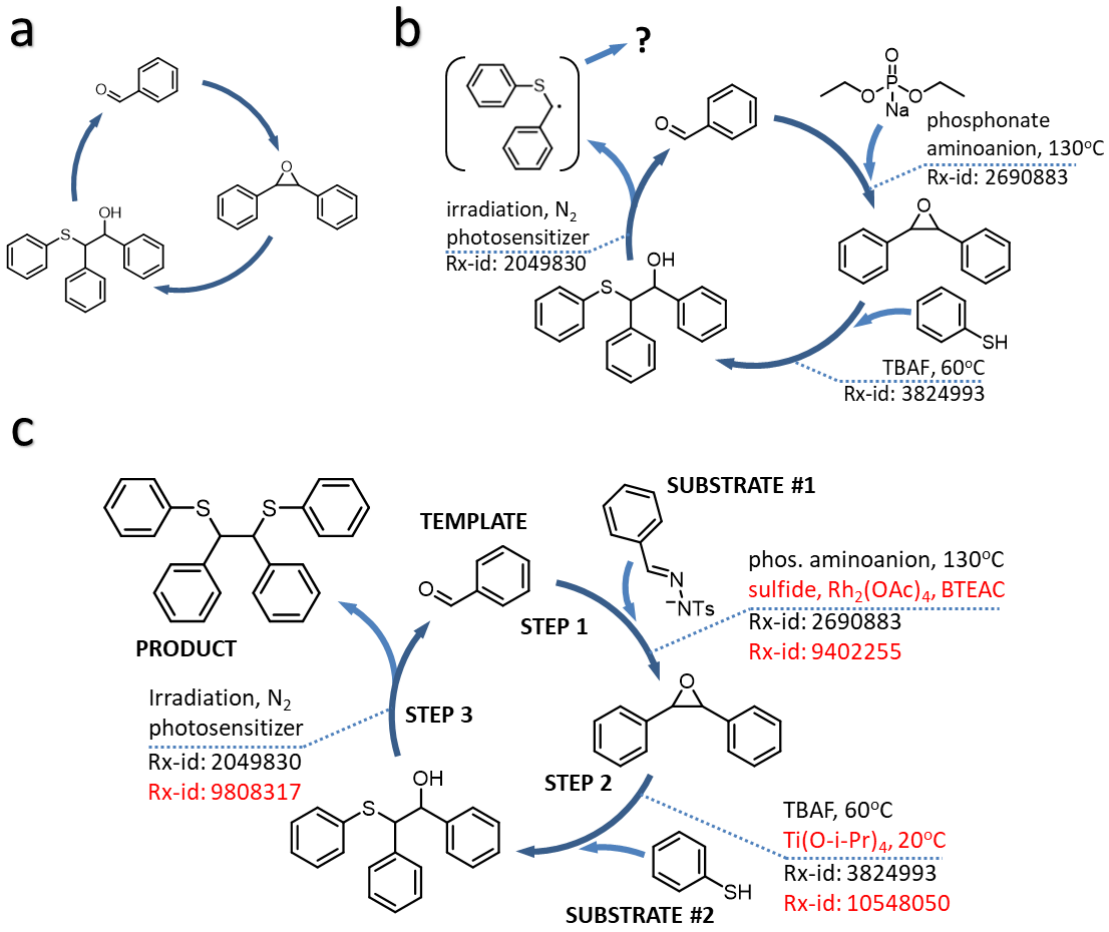


Figure 20. Three steps of the cycle identification. First, the cycle was recognized from the results page based on a promising/interesting set of intermediates, shown in (a) – the proposed transformations were plausible and the differences in molecular mass suggested that the leaving side-product was a molecule larger than each entering substrate. Upon closer inspection, Cyclorg provided rx-id numbers for each reaction involved, allowing for using Reaxys for identification of the remaining reagents. As shown in (b), the initially proposed substrates were dialkylsodiophosphonate and thiophenol, but the product was not specifically reported – only its radical intermediate. Searching the chemical literature for alternative reaction conditions resulted in a modified cycle, shown in (c). The side-product was also identified from another literature source. As reference, the rx-ids and conditions originally proposed by Cyclorg are shown in black, while the alternatives proposed according to wider chemical knowledge are shown in red. Although the cycle is conceptually interesting and chemically plausible, its side-product has no reported application – therefore it was not published outside of this thesis.

4.3. Defining the filters and search criteria

As discussed in the preceding section, application of the hard-coded filters allowed the algorithm to identify a significantly improved collection of over 18 million cycles, all of which are now stored in the Cyclorg repository. The question remained, however, how to efficiently manage such a vast number of cycles. Browsing would certainly necessitate the use of additional filters, which could be readily defined and adjusted, as shown in **Figure 21**. A few such options were already available at the time – to search by structure and molecular mass of the cycle intermediates, by a time period during which the constituent reactions were published, or by a time gap between the newest and oldest reaction and/or size of the cycle (2 to 5 steps).

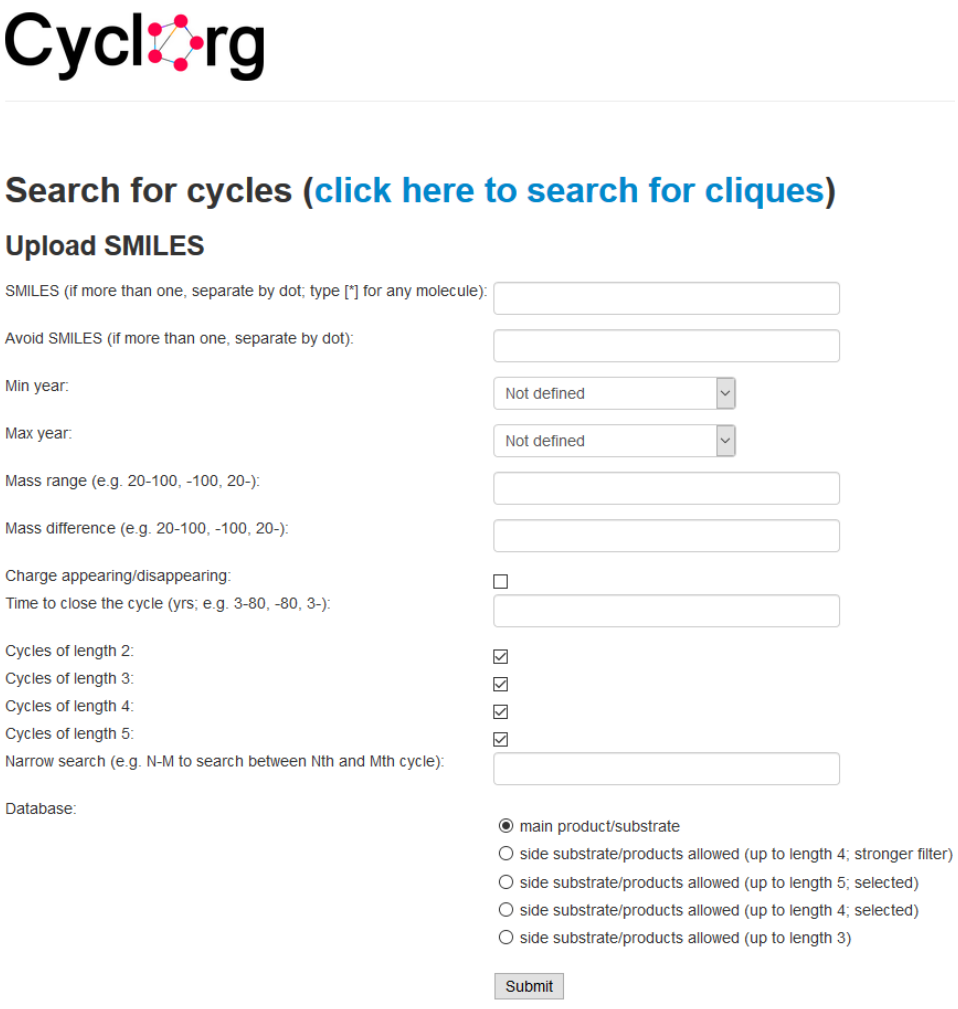
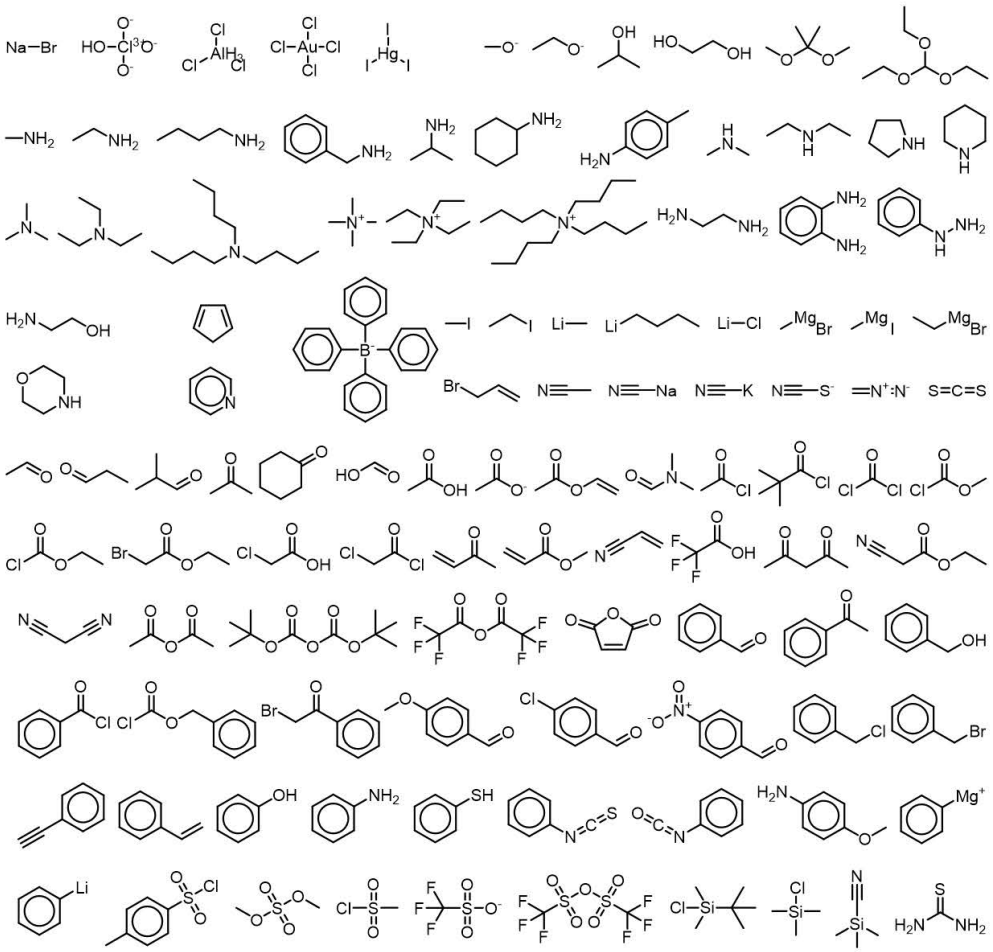
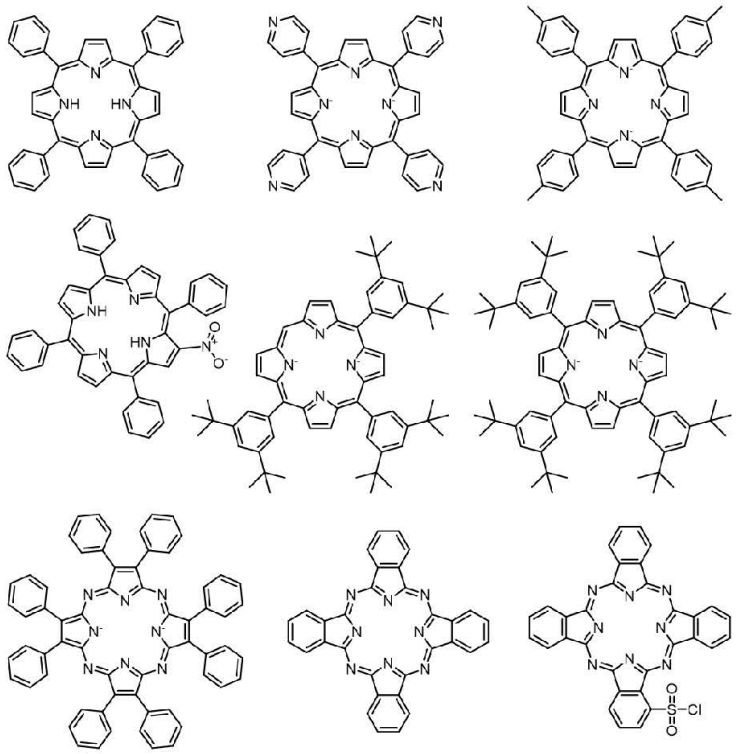


Figure 21. The UI of Cyclorg in its current form includes both the original as well as new features. The collection from which the published cycles were identified is now split and comprises the last three sub-collections available from the “Database” menu.

The “SMILES” tool allows one to search for specific chemistries. By filling the record with the SMILES string, the user defines one or more molecules to be involved in the cycles. In the initial version of Cyclorg, this query was mandatory – at least one molecule had to be specified in order to enable the search. Furthermore, only complete structures were accepted – it was not possible to include unspecified fragments, Markush structures or atom lists. These two limitations combined poorly with the available collections, reducing Cyclorg to a game of guesswork for specific molecules. The queries involving large molecules had a negligible chance of returning any results, while the queries involving small molecules produced long, recurrent lists of marginally varied cycles. The reason is that 99.5% of all generated cycles are within the NOC’s highly connected “core”, which groups approximately 225,000 useful but simple molecules out of total of ~6.3 million of the NOC members.[79] The ability to perform inclusive searches was restored by allowing the *unspecified* fragment *[*]* in SMILES. Thus, the filter is no longer mandatory (as “*[*]*” query will allow for all molecules) and it is now possible to effectively search for diverse/analogous chemistries.

While the changes to the “SMILES” made the 0.5% of cycles from outside the NOC’s “core” more accessible, the exclusion of the remaining 99.5% was enabled by addition of the complementary “Avoid SMILES” feature. By using this tool, the user can omit repeating or otherwise undesirable cycles as none of the results will involve the specified molecules. For the user’s convenience, my personally recommended list of avoidable SMILES is provided later in this section and illustrated in **Figure 22**. I do not recommend inputting *[*]* into the “Avoid SMILES”.





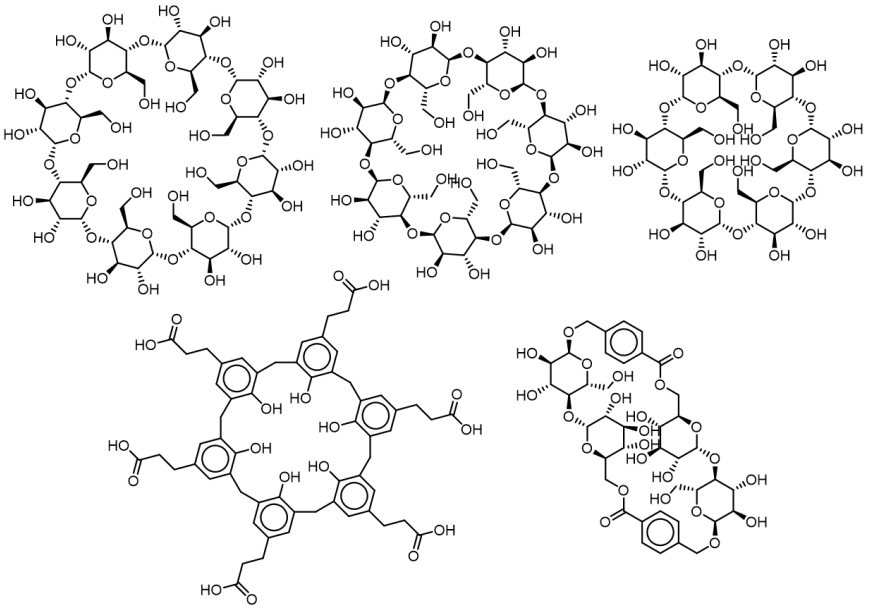


Figure 22. When searching for cycles, it is highly recommended to always avoid the presented structures. The copiable SMILES set is available in the main text. Fullerene derivatives (total of three), ammonia, hydrogen sulfide and following ions are omitted from this figure: F^-, Tb^{3+}, AlH_3^{3+}, Ga^{3+}, Ba^{2+}, Ca^{2+}, Sm^{3+}, Ni^{2+}, Na^+, SbH_8^{5+}, Sr^{2+}, Ru^{2+}, Eu^{3+}, Co^{2+}, Sn^{4+}, Cu^+, Mn^{3+}, Cu^{2+}, Pr^{3+}, Yb^{3+}, Hg^{2+}, Cs^+, Li^+, Hg^{2+}, Ag^+, Gd^{3+}, Fe^{2+}, Rb^+, Ti^{4+}, Fe^{3+}, Zn^{2+}, Pd^{2+}, Ti^{4+}, Mg^{2+}, PbH_8^{4+}, Dy^{3+}, K^+, Cr^{2+}, OH^-, Lu^{3+} and H^+.

Another way to exclude trivial cycles was provided by the “Mass range” tool, which allows for setting maximum and/or minimum molecular mass of the intermediates involved. While useful in its own right, this feature is unable to control the distribution of masses within each cycle. For example, if the allowed masses are set between 100 Da and 500 Da, it is entirely possible for the cycles returned to involve only compounds weighing less than 150 Da. This was an apparent downside as equal distribution of mass was typically found in futile cycles – on the other hand, the cycles that synthesized the sizeable side-products exhibited a significant disparity. The latter is exemplified in **Figure 20** and more examples are available in sub-section **4.4.3.**

As a means to promote synthetically useful candidates, the “Mass difference” was implemented. It allows one to impose a difference in molecular mass between the lightest and heaviest intermediates of the cycle. Setting a high minimum greatly reduces the number of futile cycles, while setting a narrow range promotes specific side-products.

Taken together, incorporation of these additional features has significantly improved the ability of Cyclorg to select truly interesting cycles – however, the technical issues caused by the sheer size of the collection still needed to be addressed. We felt that the results page load time was too long – thus, in order to keep it reasonable, Cyclorg was limited to return only the first 1,000 cycles from each search. Although necessary, this change was not without a fault. Each cycle had a discrete id number assigned to it – provided they meet the search criteria, the cycles of lower id number will always show before the ones of higher id. Not only did the set display limit result in a bias, but it also made some cycles practically inaccessible. While technically it was possible to upload the precise set of SMILES required to find the particularly elusive cycle, it would require the foreknowledge that defeats the purpose of Cyclorg. As the validation process necessitated access to every cycle, the "Narrow search" option, which filters the cycles by their id number, was added. Apart from giving access to specific cycles, it allows to "split" the collection and navigate it on a segment-by-segment basis (e.g., by performing a search for cycles numbered 0 to 1,000, then another one for cycles numbered 1,001 to 2,000 and so on). All four available collections (referred by the UI as "databases") use their own numbering – thus, if a cycle exists in two or more collections, it will have a different id in each.

Throughout the progress of the cycle validation, I managed to establish an arguably optimal set of the search criteria, which I recommend to new users. This is a baseline setup, which is supposed to compensate the NOC's inherent downsides – for more specific searches, the new criteria should be added on top of the following:

- **SMILES = [*]** (any molecule) – specifying the structure is only advisable in specialized searches,

- minimum year = 1970 – older literature is at times less credible and may result in cycles based on obsolete chemistries,

- **mass range = at least "100- " and up to "600- "** – very small compounds, like alkyl alcohols, return overwhelming numbers of the repetitive cycles based on trans-esterification. Very large compounds – typically porphyrins, peptides and polysugars – return futile cycles based on functional group transformations or polymerization/hydrolysis. The existence of purposeful

cycles featuring such complex compounds is plausible – however, it would require more steps than current maximum of 5,

- **mass difference = at least "200- "** – this parameter is essential in finding cycles that produce sizeable by-products – the minimum value is vital, but maximum is up to the user's preferences,

- **time to close cycle = at least "1-"** – this excludes cycles that have been reported entirely in one publication. The minimum might need to be increased for a candidate that multiple independent groups have worked on, as it is not uncommon for a synthetic pathway to be published in multiple articles over the span of 2-5 years,

- **narrow search = "0-250,000" then "250,000-500,000" etc.** – or if Cyclorg returns 1,000 examples, check the id of the last cycle in the results list to see where to begin the next search,

- **"side substrate/product" databases** are strongly recommended as they allow smaller reactants to appear as intermediates and larger ones to appear as by-products,

- **avoid Smiles =** C(CBr)(c1ccccc1)=O.C(Cc1cc2c(O)c(c1)Cc1cc(cc(c1O)Cc1cc(cc(c1O)Cc1cc(cc(c1O)Cc1cc(cc(c1O)Cc1cc(cc(c1O)C2)CCC(O)=O)CCC(O)=O)CCC(O)=O)CCC(O)=O)CCC(O)=O)C(O)=O.OC(C)C.c1cc(ccc1)C(Cl)=O.C(=O)(Cl)C.OC[C@@H]1[C@H]2O[C@H]3O[C@@H]([C@@H](O[C@H]4O[C@@H]([C@@H](O[C@H]5O[C@@H]([C@@H](O[C@H]6O[C@@H]([C@@H](O[C@H]7O[C@@H]([C@@H](O[C@@H](O1)[C@H](O)[C@H]2O)[C@H](O)[C@H]7O)CO)[C@H](O)[C@H]6O)CO)[C@H](O)[C@H]5O)CO)[C@H](O)[C@H]4O)CO)[C@H](O)[C@H]3O)CO.n1ccccc1.c12c(c3ccc(c(c4nc(c(-c5ccccc5)c5ccc(c(c(n1)cc2)-c1ccccc1)[nH]5)c([N+](=O)[O-])c4)-c1ccccc1)[nH]3)-c1ccccc1.c12c3c4c5c6c7c8c9c%10c%11c%12c%13c%14c%15c%11c(c1c%15c1c%11c%15c%16c%17c%18c%19c%15c%15c%20c%21c(c5c5c%21c%21c(c%19%20)c%19c%20C%22%23c(c%18%19)c%18-c%19c%24c%25c%26c(C%22%24C%23(P(OCC)(OCC)=O)P(OCC)(OCC)=O)c-%22c(c(c75)c%21c%20-%22)c8c%26c9c%12c%25c%13c%19c(c%14c1%16)c%17%18)c3c%15c2%11)c4c6%10.C(=O)(Cl)Cl.C=CC(C)=O.c1(ccccc1)-c1c2nc(c1-c1ccccc1)nc1c(c(-c3ccccc3)c(nc3nc(c(-c4ccccc4)c3-c3ccccc3)nc3c(c(-c4ccccc4)c(n2)[n-]3)-c2ccccc2)[n-]1)-c1ccccc1.c1ccc(cc1)N.CCOC(CBr)=O.O1[C@H]2O[C@H](COC(c3ccc(CO[C@@H]4[C@H](O)[C@@H](O)[C@H](O[C@H]5O[C@H](COC(c6ccc(CO[C@@H]7[C@H](O)[C@@H](O)[C@H]1[C@H](O7)CO)cc6)=O)[C@@H](O)[C@H](O)[C@H]5O)[C@H](O4)CO)cc3)=O)[C@@H](O)[C@H](O)[C@H]2O.Cc1ccc(cc1)-c1c2ccc(c(c3ccc(c(c4ccc(c(c5ccc1[n-]5)-c1ccc(C)cc1)n4)-c1ccc(C)cc1)[n-]3)-c1ccc(C)cc1)n2.[Cl+3](O)([O-])([O-])[O-].N#C[Si](C)(C)C.OC(CCl)=O.N1CCOCC1.IC.[PbH8+4].C(Br)c1ccccc1.O(C(=O)C)C=C.O=C=Nc1ccccc1.N(c1ccccc1)=C=S.c1cccc2c3nc4c5cccc(c5c(nc5c6c(c(n5)nc5c7c(c(nc(n3)c12)n5)cccc7)cccc6)n4)S(Cl)(=O)=O.C(Cl)c1ccccc1.c12ccc(c(c3ccc(cc4ccc([n-]4)c(-c4cc(cc(c4)C(C)(C)C)C(C)(C)C)c4ccc(c1-c1cc(cc(C(C)(C)C)c1)C(C)(C)C)n4)n3)-c1cc(C(C)(C)C)cc(C(C)(C)C)c1)[n-]2.c1cc2c(c3[n-]c(c(-c4ccncc4)c4ccc(c(c5[n-]c(c(-c6ccncc6)c1n2)cc5)-c1ccncc1)n4)cc3)-c1ccncc1.CC(O)=O.[Lu+3].c1c(cccc1)C#C.NNc1ccccc1.[OH-].S(=O)(=O)(OC)OC.CC(C)(OC)OC.CC[Mg]Br.c1c(cccc1)COC(=O)Cl.c1ccc(cc1)CN.c1(ccc(C=O)cc1)OC.C1CC(CCC1)N.C(C(=O)OC)=C.[N-]=[N+]=C.c1([B-](c2ccccc2)(c2ccccc2)c2ccccc2)ccccc1.[AlH3](Cl)(Cl)Cl.C(OC(Cl)=O)C.c

1cc(C=O)ccc1[N+](=O)[O-].[K]C#N.C(C)CCN.C(C)N(CC)CC.c1cc(Cl)ccc1C=O.C1C(CCCC1)=O.ClS(C)(=O)=O.N(C)C.C(C)(C)(C)OC(=O)OC(=O)OC(C)(C)C.O[C@H]1[C@@H]2O[C@@H]([C@H]([C@@H]1O)O[C@H]1O[C@@H]([C@H]([C@@H]([C@H]1O)O)O[C@H]1O[C@@H]([C@H]([C@@H]([C@H]1O)O)O[C@H]1O[C@@H]([C@H]([C@@H]([C@H]1O)O)O[C@H]1O[C@@H]([C@H]([C@@H]([C@H]1O)O)O[C@H]1O[C@@H]([C@H]([C@@H]([C@H]1O)O)O2)CO)CO)CO)CO)CO)CO.c1ccc(CO)cc1.[Fe+3].[Fe+3].C[Si](C)(C)Cl.C[Mg]Br.C([S-])#N.Nc1c(N)cccc1.C(C)OC(CC(=O)OCC)=O.C(OC(CC#N)=O)C.C(=O)C.O=Cc1ccccc1.C(=O)(CCl)Cl.C(F)(F)(F)S(=O)(=O)OS(C(F)(F)F)(=O)=O.NC(N)=S.C=Cc1ccccc1.[Gd+3].C(C)(C)C=O.c12c(cccc1)c1nc3c4c(cccc4)c(nc4c5c(cccc5)c(nc5c6c(cccc6)c(nc2n1)n5)n4)n3.ClC(=O)C(C)(C)C.c1ccc(cc1)[Mg+].[Na]C#N.COC(=O)Cl.C(CC)=O.[Zn++].[Zn++].[Zn++].[Zn++].C(=S)=S.c1cc(ccc1OC)N.C(F)(F)(F)C(=O)OC(C(F)(F)F)=O.C(=C)C#N.C(C[Li])CC.[O-]CC.C([O-])(C)=O.Cc1ccc(N)cc1.C(CC)CN(CCCC)CCCC.I[Hg](I)I.[C@@H]12O[C@H]3O[C@@H]([C@@H](O[C@H]4O[C@@H]([C@@H](O[C@H]5O[C@@H]([C@@H](O[C@H]6O[C@@H]([C@@H](O[C@H]7O[C@@H]([C@@H](O[C@H]8O[C@@H]([C@@H](O[C@H]9O[C@@H]([C@@H](O[C@@H]1CO)[C@@H]([C@H]2O)O)[C@@H]([C@H]9O)O)CO)[C@@H]([C@H]8O)O)CO)[C@@H]([C@H]7O)O)CO)[C@@H]([C@H]6O)O)CO)[C@@H]([C@H]5O)O)CO)[C@@H]([C@H]4O)O)CO)[C@@H]([C@H]3O)O)CO.n1c2ccc1c(-c1cc(C(C)(C)C)cc(c1)C(C)(C)C)c1ccc([n-]1)c(-c1cc(C(C)(C)C)cc(c1)C(C)(C)C)c1nc(cc1)c(-c1cc(C(C)(C)C)cc(c1)C(C)(C)C)c1ccc([n-]1)c2-c1cc(C(C)(C)C)cc(c1)C(C)(C)C.C1NCCCC1.CC(C)=O.S(c1ccc(cc1)C)(=O)(=O)Cl.[Dy+3].FC(F)(F)C(O)=O.c1(ccccc1)-c1c2nc(cc2)c(-c2ccccc2)c2ccc(c(-c3ccccc3)c3nc(cc3)c(-c3ccccc3)c3ccc1[nH]3)[nH]2.[Mn+3].[Cu++].[Cu++].[Cu++].[Cu++].[Cu++].C(C=C)Br.C(S(=O)(=O)[O-])(F)(F)F.[N+](CC)(CC)(CC)CC.c1ccc(cc1)[Li].O(C(OCC)OCC)CC.CCl.[K+].c1(ccccc1)O.Br[Na].Cl[Li].C1=CCC=C1.O=CN(C)C.c1ccc(S)cc1.O(C(C)=O)C(C)=O.[Ca++].C1=CC(=O)OC1=O.[Ga+3].[Ba++].C1NCCC1.CC(c1ccccc1)=O.C[N+](C)(C)C.C[Si](C)(C)(C)C)Cl.N#CCC#N.[AlH3+3].C(C)NCC.Cl[Au](Cl)(Cl)Cl.[Sm+3].[Ni++].[Ni++].[Ni++].[Ni++].C[Mg]I.O=CO.[Na+].[SbH8+5].NCCO.C(C)#N.NCC.[Li]C.C(CN)N.CN(C)C.[Cr].C(C)(C)N.[Yb+3].CCCC[N+](CCCC)(CCCC)CCCC.[Eu+3].[Co++].[Co++].[Sn+4].[Sn+4].[Sn+4].[Cu+].C[O-].C(CO)O.NC.[Fe++].[Fe++].[Fe++].S.[H+].[Pd++].[Rb+].[Fe+4].[Hg++].[Hg++].[Ti+4].[Ti+4].[Mg++].C(C(=O)C)C(=O)C.[Ag+].[Ag+].[Cs+].[Li+].[Tb+3].c12c3c4c5c1c1c6c7c2c2c8c3c3c9c4c4c%10c5c5c1c1c6c6c%11c7c2c2c7c8c3c3c8c9c4c4c9c%10c5c5c1c1c6c6c%11c2c2c7c3c3c8c4c4c9c5c1c1c6c2c3c41.[Sr++].[Sr++].[Ru++].[Ru++].N.[F-].[Pr+3]
– see **Figure 22** for drawn structures.

In order to further improve the quality of the “side substrate/product” databases (i.e., allowing smaller reagents to appear in the vertices of cycle graphs) in the public version of Cyclorg, I have defined an improved set of hard-coded filters. Based on my unique experience and on the list of substances appearing in the greatest number of cycles (see **Figure 23**), I have defined the new filter as a set of ten rules:

– exclude a cycle if any intermediate comprises three non-hydrogen atoms or less and none of them is sulfur, selenium or phosphorus,

– exclude a cycle if any intermediate comprises O=COC=O or [*][C](=[O])[O][C]([*])=[O] motif and less than 16 non-hydrogen atoms,

– exclude a cycle if any intermediate comprises ClC=O, BrC=O, IC=O or [*][C]([Cl,Br,I])=O and less than 9 non-hydrogen atoms,

– exclude a cycle if any intermediate comprises CN, CNC, Nc1ccccc1, [CX4,c][NH2], [CX4,c][NH][CX4,c] or [CX4][N]([CX4])[CX4], no other heteroatoms and less than 9 non-hydrogen atoms, (bond order matters, so C#N is acceptable)

– exclude a cycle if any intermediate comprises CCl, CBr, CI, CF or [CX4][Cl,Br,I,F], no other heteroatoms and less than 13 non-hydrogen atoms,

– exclude a cycle if any intermediate comprises solely oxygen, carbon and hydrogen atoms and only single non-aromatic bonds (i.e., comprises no multiple bonds), regardless of the atom count or mass,

– exclude a cycle if any intermediate comprises [Si][Cl] and only carbon as non-hydrogen atoms, (accordingly, [Cl][Si][Cl] is acceptable),

– exclude a cycle if any intermediate comprises c1ccccc1, less than 9 non-hydrogen atoms and no selenium or phosphorus atoms,

– exclude a cycle if any intermediate comprises the porphyrin ring N1C2=CC=C1\C=C1\C=CC(\C=C3/N\C(\C=C3)=C/C3=N/C(/C=C3)=C\2)=N1, regardless of other factors,

– exclude a cycle if any intermediate comprises C1C2CC3CC1CC(C2)C3 (the adamantane scaffold) and less than 20 non-hydrogen atoms,

– exclude a cycle if any intermediate is one of the following: [H+], [Li+], N, [Na+], [Mg++], [AlH3+3], S, [K+], [Ca++], [Cr], [Mn+3], [Fe++], [Fe+3], [Fe+4], [Co++], [Ni++], [Cu++], [Zn++], [Ga+3], [Rb+], [Sr++], [Ru++], [Ag+], [Sn+4], [SbH8+5], [Cs+], [Ba++], [Sm+3], [Eu+3], [Tb+3], [Dy+3], [Yb+3], [Lu+3], [Hg++], [PbH8+4], [Na]Br, Cl[AlH3](Cl)Cl, I[Hg](I)I, C[N+](C)(C)C, O[Cl+3]([O-])([O-])[O-], Cl[Au](Cl)(Cl)Cl or OC(=O)C(F)(F)F.

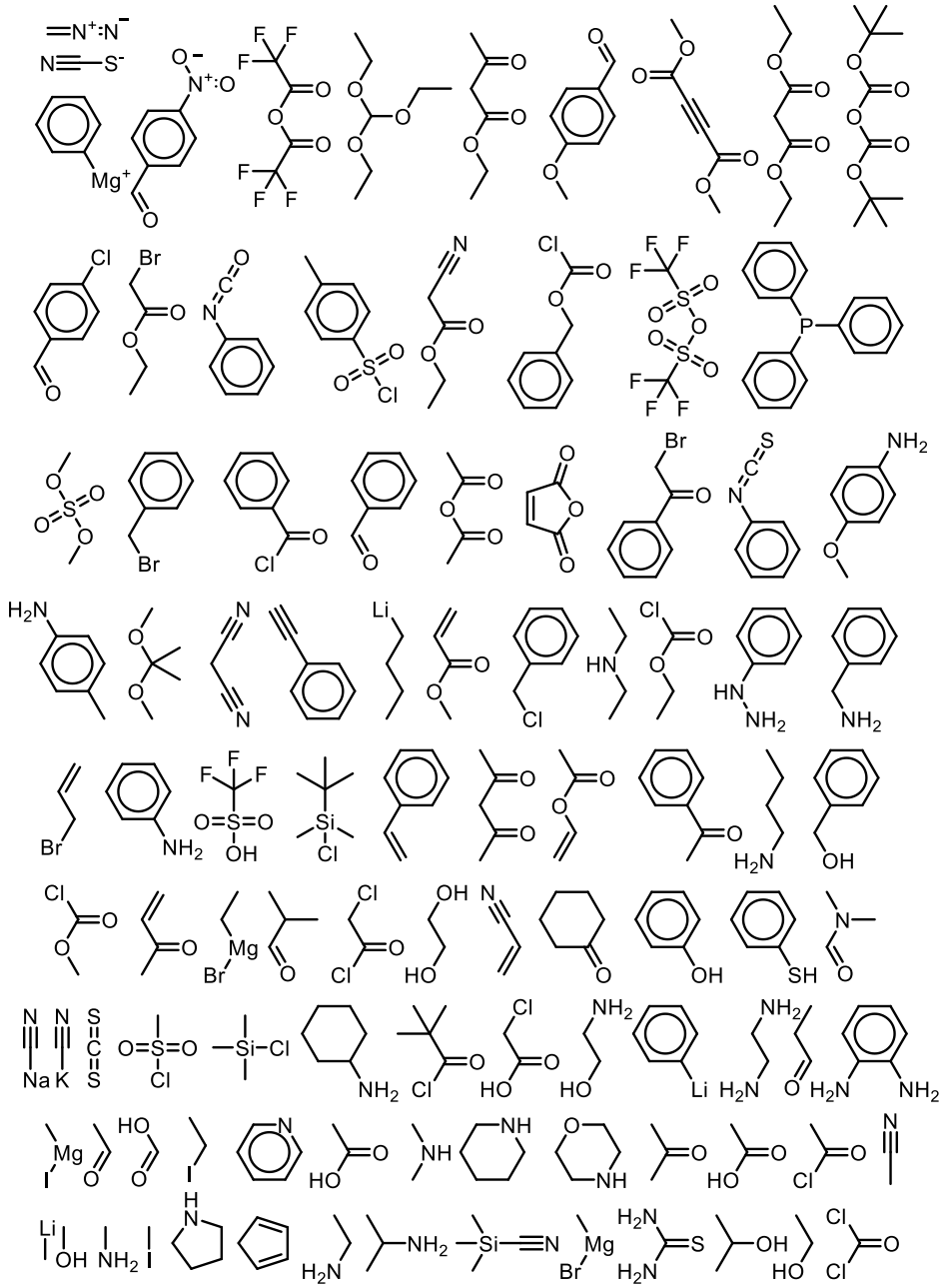


Figure 23. The molecules that appeared in the greatest number of cycles. Although not specifically forbidden, they were useful for defining the advanced filters. The candidates that involved these compounds varied only marginally and appeared in vast numbers that impeded the searches, effectively acting as the “noise” present in the results.

4.4. Identified cycle candidates

This section provides detailed descriptions of some of the most interesting cycles identified from the Cyclorg repositories. From this collection, some of the most illustrative examples were published in our *Angewandte Chemie*[19] paper – these examples combine usefulness, diverse chemistries and, in most cases, works of independent chemists. However, they are not supposed to work one-pot unless specified. A number of unpublished candidates is also described – as less than optimal, these examples mostly provide insights into the methodology applied for their discovery.

The following parts describe some main types of cycles: those featuring auto-amplification (sub-section **4.4.1.**), enzyme-like biomimesis (**4.4.2.**), progressive assembly (**4.4.3.**), close relation to other cycles (**4.4.4.**) and asymmetric catalysis (**4.4.5.**). Various sub-types, such as triphenylphine-based (**3.4.6**) or metabolically biomimetic cycles (**3.4.7**) and specific examples like the “Moebius strip-like” cycle (**4.4.8.**), are also shown in the remaining sections.

4.4.1. Auto-amplification cycles

The cycles featuring auto-amplification of their components are arguably most exciting and relevant to systems chemistry. As discussed in Section **3.3**, an initial intermediate uses available substrates and cofactors to produce its own copy over a number of chemical reactions without being ultimately consumed. The examples presented demonstrate how cycle-assembling tools could be used to approach, bottom-up, the problem of the automatic discovery of auto-amplifying chemical systems.

Figure 24 illustrates one useful auto-amplifying cycle, which has been published in our *Angewandte Chemie*[19] paper. In this cycle, a popular chiral auxiliary, (S)-4-benzyl-2-oxazolidinone (marked with the blue star), enters the cycle to participate in an enantiomerically pure synthesis of its own copy. The cycle starts with acylation of the chiral auxiliary with 3-phenylpropionic acid chloride (90% reported yield) followed by diastereoselective α-azidation by trisyl azide (91%). The further cleavage of the oxazolidinone auxiliary produces a chiral α-azido acid (100%, quantitative), while also recovering the chiral auxiliary itself (the dashed arrow across the cycle in **Figure 24**). The azido acid is subsequently reduced to (S)-phenylalanine (86%). The final step is the recreation of (S)-4-benzyl-2-oxazolidinone via one-pot reduction acylation of the amino acid (78%). All in all, the chiral auxiliary is created “along the cycle” and via one connection “across it” — importantly, the literature-reported yields of individual reactions are such that for every starting gram of this auxiliary that enters the cycle, 1.37 grams are generated after the cycle is executed. It is therefore true auto-amplification. Although not amenable to one-pot operation, an interesting caveat of this cycle is that if a different acid chloride is used at input, the three-membered “inner” cycle (i.e., closing along the dashed arrow) can produce different chiral azido acids, which can ultimately give different amino acid derivatives.

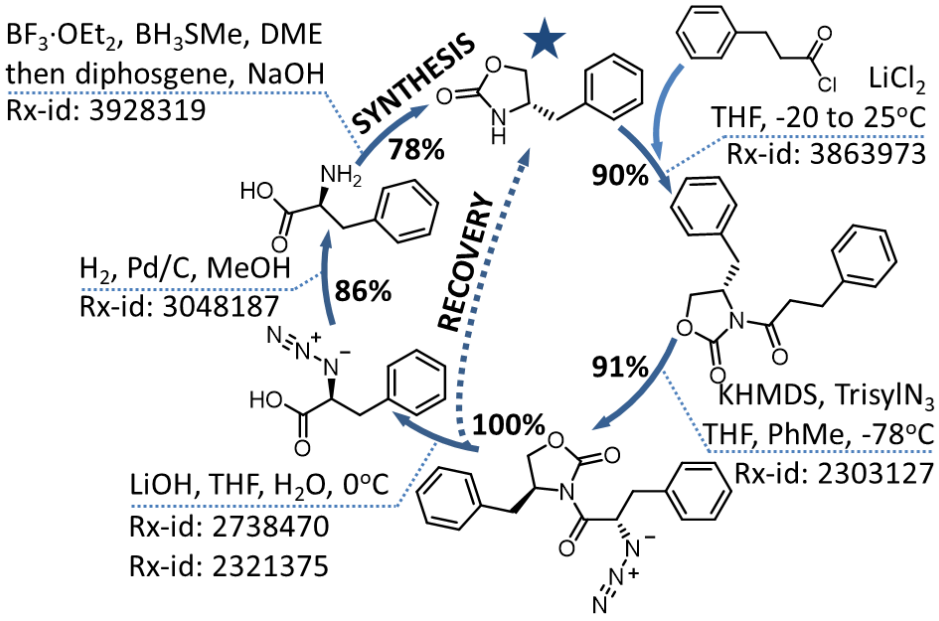


Figure 24. The "auto-amplifying" cycle in which a popular chiral auxiliary (marked with the star) is regenerated over the cycle, but it also makes its own additional copy as a side product of one of the reactions.

In contrast to *in vivo* self-replicating systems, yield becomes of significant concern in chemical-organic cycles. Living cells comprise dynamic systems, where reactions proceed continuously – under mild, aqueous conditions and with the involvement of highly selective enzymes. Organic chemists, on the other hand, typically perform individual reactions separately, then follow with quench, work-up and product purification. As the cycles are assembled from organic chemistry reports, each of their steps is characterized by a specific yield – together they comprise the cycle total yield, which, if low enough, may nullify the otherwise apparent auto-amplifying effect.

A specific cycle whose low reported yield prevents the true auto-amplification, is shown in **Figure 25**. This time a different Evans reagent – (4R,5S)-(+)-4-Methyl-5-phenyl-2-oxazolidinone, marked with the blue star – is being regenerated. It enters the cycle through acylation with propionyl chloride (90% reported yield), followed by the asymmetric aldol reaction with benzaldehyde (91%). The further cleavage of the oxazolidinone auxiliary produces chiral methyl (2R,3R)-3-hydroxy-2-methyl-3-phenylpropanoate (66%), but also recovers the auxiliary, as shown with the dashed arrow. The produced ester is hydrolyzed (72%) and transformed into a copy of starting chiral auxiliary (100%). Although recovery of the auxiliary during the cleavage was not specifically reported,

the wider chemical literature allows to predict this as plausible – the precise yield is unknown. Assuming that the yield of oxazolidinone recovery would be the same as of propanoate (66%), the total yield of the Evans reagent would convey the net loss every time the cycle is completed (0.84g generated for every starting gram).

There are no other literature reports on the two lowest-yielding steps – cleavage and hydrolysis. The contemporary sources propose better conditions, but only for reactions with diastereoisomers of original compounds. In such cases, the yields were 93% and 90%, respectively (as shown with red color in **Figure 13**).[80,81] These results suggest that the reported conditions were sub-optimal since changing the chiral configuration on such a small molecule would, arguably, have little to no effect on the non-stereospecific reactions, such as ester hydrolysis. In such case, the expected yield would be 1.46 grams per one starting gram – experimental verification is required.

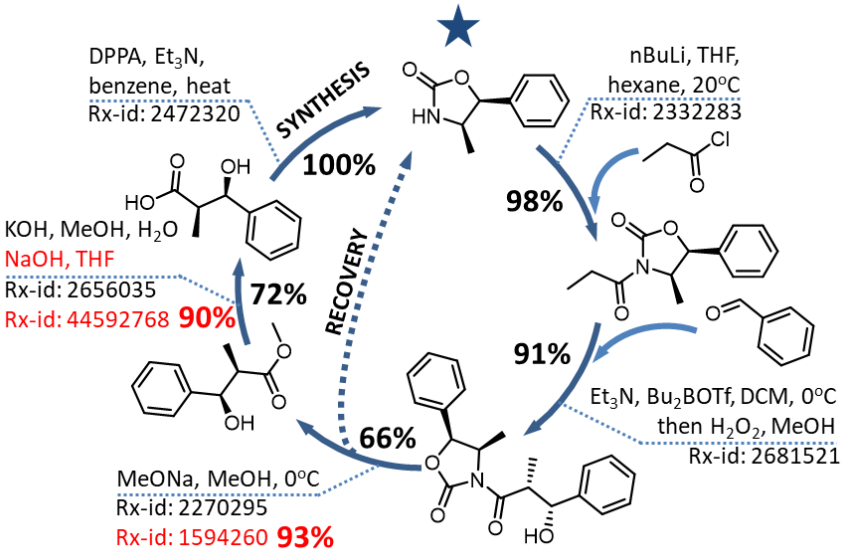


Figure 25. The example of a cycle candidate in which some lower-yielding steps prevent overall auto-amplification. As variation on the example shown in **Figure 24**, it involves auto-replication of another Evans reagent (marked with the star). The black rx-ids provide original literature entries, while the red rx-ids provide additional literature suggesting that yields might be improved up to numbers shown in red, restoring the auto-amplification effect.

Both examples shown in **Figures 24-25** exhibit similar architectures (i.e., two cycles joined by a shared sequence of reactions and a starting reagent), arguably forecasting some potential methods of the automatic auto-amplifying cycles discovery (e.g. an improved search algorithm, which identifies the network architectures typical for such cycles). An additional insight into such methods could be derived from the negative results of Cyclorg. **Figure 26** shows an example of the false-positive "auto-amplification". Here, (2S,3S,4R,6R)-3-hydroxy-2,4,6-trimethyloctanoic acid (marked with the blue star) reacts as lactone with another substrate, i.e., (3R,4S,5R,7R)-3,5,7-trimethyl-1-nonen-4-ol, to form a product that upon further transformation hydrolyzes into two equivalents of the starting reagent – an original and a copy. It is, however, not true auto-amplification. Apart from the poor yields, the involvement of the starting compound in its reproduction is useless – it is far more efficient to transform the side-substrate from step 2 directly into the final product. The first three steps were initially published for the synthesis of bourgeanic acid (a product of the 3rd step) and the final hydrolysis was reported independently as a crystallographic study. Without understanding of actual chemistry, the algorithm could not recognize this example as false-positive, showing that the adequate network architecture is insufficient on its own to guarantee auto-amplification.

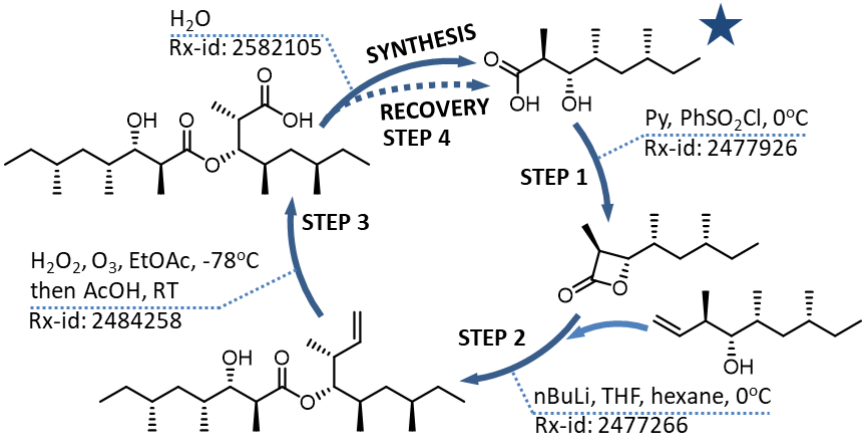


Figure 26. Despite having the requisite network architecture, this cycle does not feature auto-amplification. The starting reagent, i.e. ,(2S,3S,4R,6R)-3-hydroxy-2,4,6-trimethyloctanoic acid, marked with the blue star, is incorporated into a larger compound, which later transforms back into the original substrate and its copy.

4.4.2. Enzyme-like rearrangement cycles

Majority of cycles are non-autocatalytic though many are still biomimetic and relevant to systems chemistry. A close analogy to enzymatic catalysis was observed in a group of the cycles based on intramolecular reactions. In those cycles, the starting intermediate binds all the substrates one-by-one, bringing them "close together" within one molecular skeleton to promote a specific reaction. Once this reaction (a rearrangement or cyclization) is complete, the newly-formed compound is cleaved into a product and intermediate, which is transformed back into its starting form. Such regenerating molecules effectively function as "molecular templates" for synthesized products – this is how they will be referred to in the remaining parts of this work.

An example of a template-based cycle is shown in **Figure 27**. Here, the dimer of the template, 2,2'-dithiobis(N-methylaniline), undergoes acylation with two equivalents of 11-bromoundecanoyl chloride to form a dimer of two substrate-template adducts. This dimer is split into two by the reductive cleavage of the disulfide bond and each resulting "half" undergoes independent internal cyclization, giving two equivalents of the cyclic lactam intermediate. The third step comprises α-methylation of this lactam, followed by oxidation of its sulfide group. Next is the most important step of the cycle, where its product is formed on the template during intramolecular cyclization, resulting in a large cycloundecanon compound. In the last step, the product is cleaved off as 11-methylcycloundec-2-en-1-one, leaving the template that immediately dimerizes into its initial form of 2,2'-dithiobis(N-methylaniline). This entire cycle was developed by the Ohtsuka group over the course of 4 years – the work published in 1979 describes the synthesis of cyclic ketones[82] and the report from 1983 extends the method to unsaturated ketones[83]. The retrieval of the disulfide during the last step was also reported for the first time in 1983, thus closing the cycle.

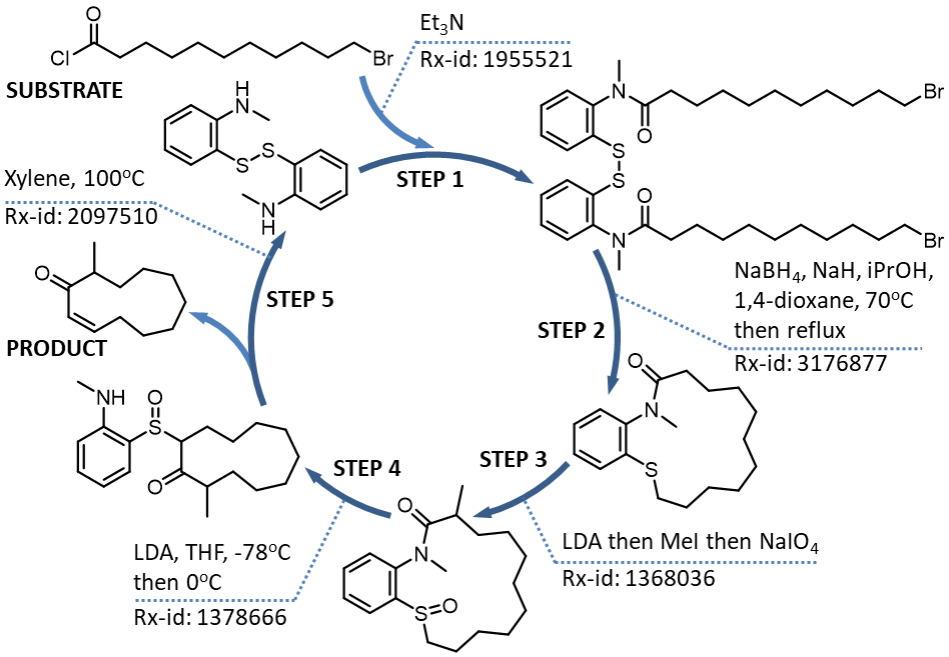


Figure 27. An example of the cycle based on intramolecular cyclization. It produces 11-methylcycloundec-2-en-1-one from 11-bromoundecanoyl chloride and methyl iodide with o-methylaminothiophenol acting as a molecular template in an enzyme-like manner by keeping both terminal reaction sites close during cyclization.

Because it was assembled from the works of just one research group, we have not published the Ohtsuka's cycle. Another candidate has been chosen instead, as shown in **Figure 28**. This time it took independent research from across the span of 157 years to close the cycle. The cycle begins with the succinic acid template being transformed into an anhydride, then reacting with hydrazine and subsequently with 4-methoxy-benzoyl chloride. The resulting compound undergoes the rearrangement that produces 2,5-bis(4-methoxyphenyl)-1,3,4-oxadiazole and regenerates the template acid. This is a very practical example, as the diaryl-oxadiazole scaffold produced by this cycle is of recent interest in molecular electronics due to its electron-transporting and hole-blocking properties.[84,85] More importantly, the cycle could potentially be operated one-pot (in acetonitrile and changing phosphorous pentoxide to DCC) but only once due to the water quench in the last step.

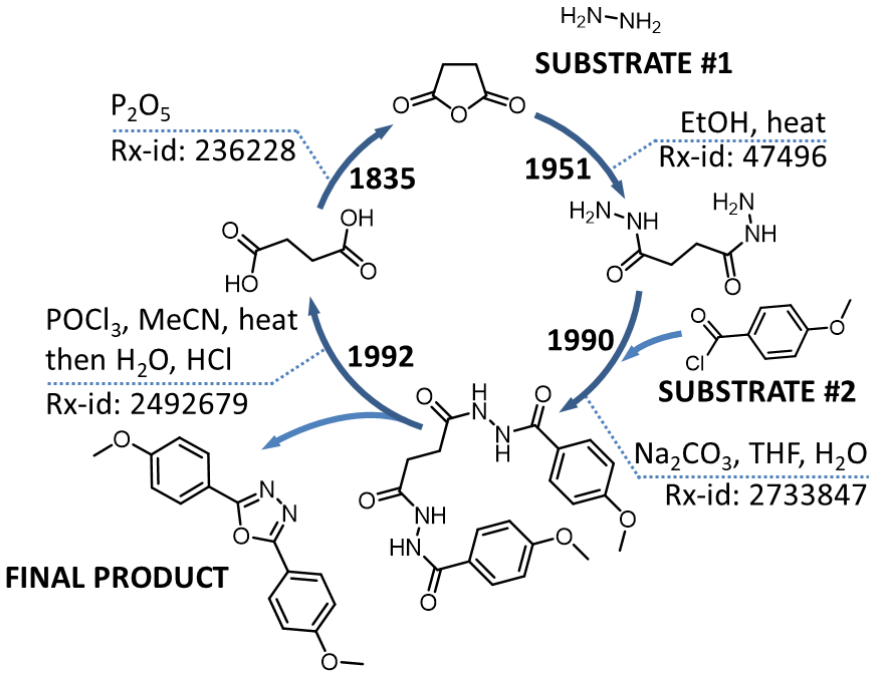


Figure 28. The cycle that took 157 years to complete. This cycle was opened in 1835 with the report of synthesis of succinic anhydride. It was closed only in 1992 with a publication describing the rearrangement that produces 2,5-diaryl-1,3,4-oxadiazoles and succinic acid.

Another cycle that we published in our *Angewandte Chemie* paper[19] is remarkably versatile as it can be extended to a broader scope of substrates than the one shown in **Figure 29**.[86] In this example, 4-chloro-benzoyl chloride and 3-methyl-2-buten-1-olsubstrates are successively added onto the phenylglycine template to give (4-chloro-benzoylamino)-phenyl-acetic acid 3-methyl-but-2-enyl ester that then rearranges (by the Claisen and then the Cope types of rearrangement) into 2-(4-chlorophenyl)-2-(3,3-dimethylallyl)-4-phenyl-5(2H)-oxazolone, from which the 1-(p-chlorophenyl)-4-methyl-3-penten-1-one product is then released. This product is used in the synthesis of derivatives of cyano-featured dihydroisoxazoles known to exhibit antibacterial activity[87] and of gem-bisprenyl-based building blocks[88] of natural product analogs.

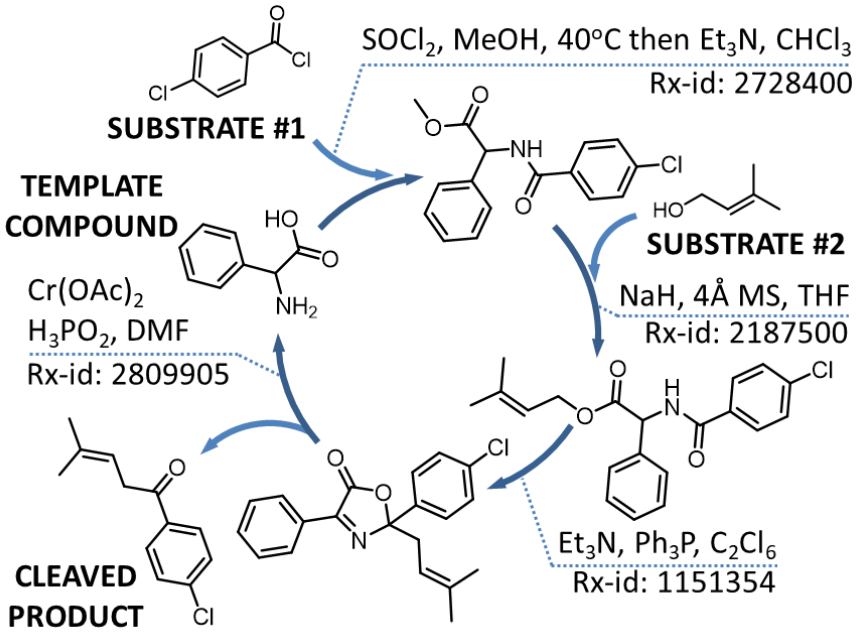


Figure 29. An example of a rearrangement-based cycle. The reactions building up mass are followed by two consecutive rearrangements (of the Claisen and then the Cope types) and then cleavage/release of 1-(p-chlorophenyl)-4-methyl-3-penten-1-one as the cycle product.

4.4.3. Template-based build-up cycles

A template compound is a frequent feature of chemical cycles. However, not all such cycles are based on intramolecular reactions. In many cases, the product is being formed gradually on the template, whilst the substrates are being added one-by-one, as opposed to one constructive event of rearrangement or cyclization. This type will be referred to as the "build-up" cycle.

The example that we published in our paper,[19] shown in **Figure 30**, best illustrates this feature whereby a complex structure is progressively being assembled on a simple template (here, dichlorodimethylsilane) to be finally cleaved away as the cycle product, in the process regenerating the template itself. Starting from the dichlorodimethylsilane, 7-silanorbornadiene ring formation and subsequent [2+4] cycloaddition yield 1,2-bis(trifluoromethyl)-3,4,5,6-tetraphenylbenzene, which is then photolyzed, producing the hexasubstituted benzene ring and regenerating dichloromethylsilane.

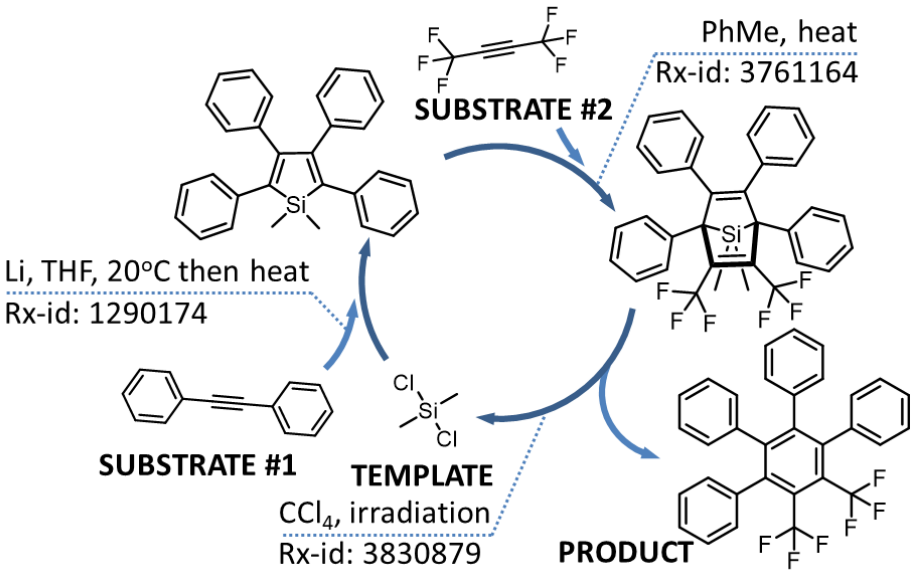


Figure 30. A "build-up" cycle in which a larger structure is progressively built on the dichlorodimethylsilane "template"— after all the substrates have been incorporated, the proper product ultimately detaches, regenerating the template.

In 1983, P. Molina group published[89] the reaction between 1-amino-4,6-diphenyl-2-pyridone and methyl(p-chlorophenyl)ketone as the first step in the synthesis of pyrido-1,3,4-oxadiazine derivatives. Just one year later,[90] the same group used the product of this reaction as a starting

material for the syntheses of pyrazoles and isoxazoles. This resulted in an example of the plausible build-up cycle that was completed within just one year (see **Figure 31**), as opposed to 157 years required to close the cycle from **Figure 28**. The cycle itself could be used to synthesize different compounds – one possible variation was already present in the Cyclorg depository, as shown in **Figure 32**.

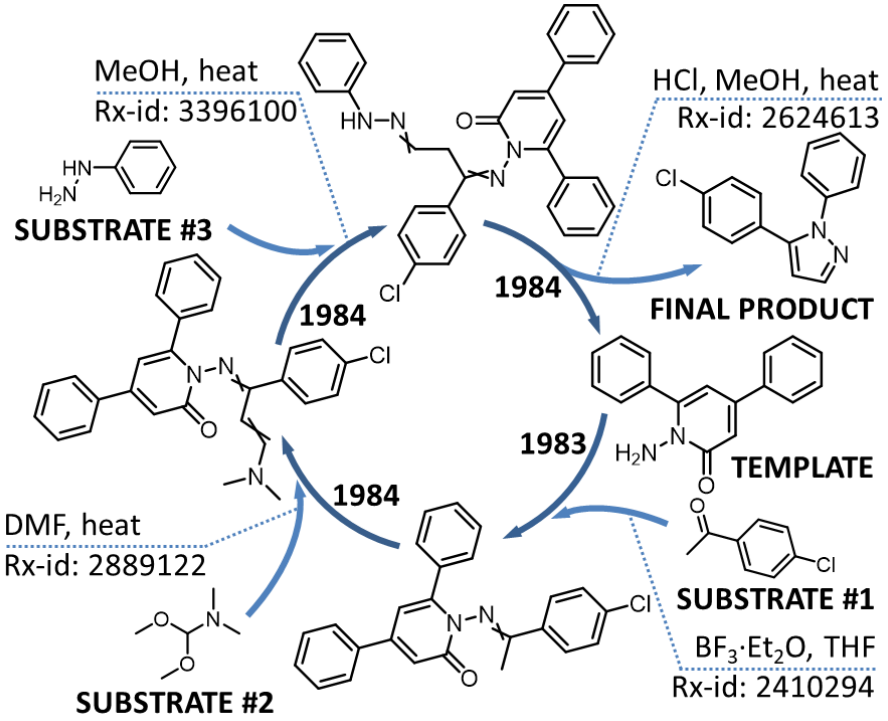


Figure 31. An example of the cycle closed by the Molina group within 1 year, but described in separate reports. This cycle could potentially be operated continuously one-pot by unifying a solvent to methanol and changing the Lewis acid to a Bronsted acid.

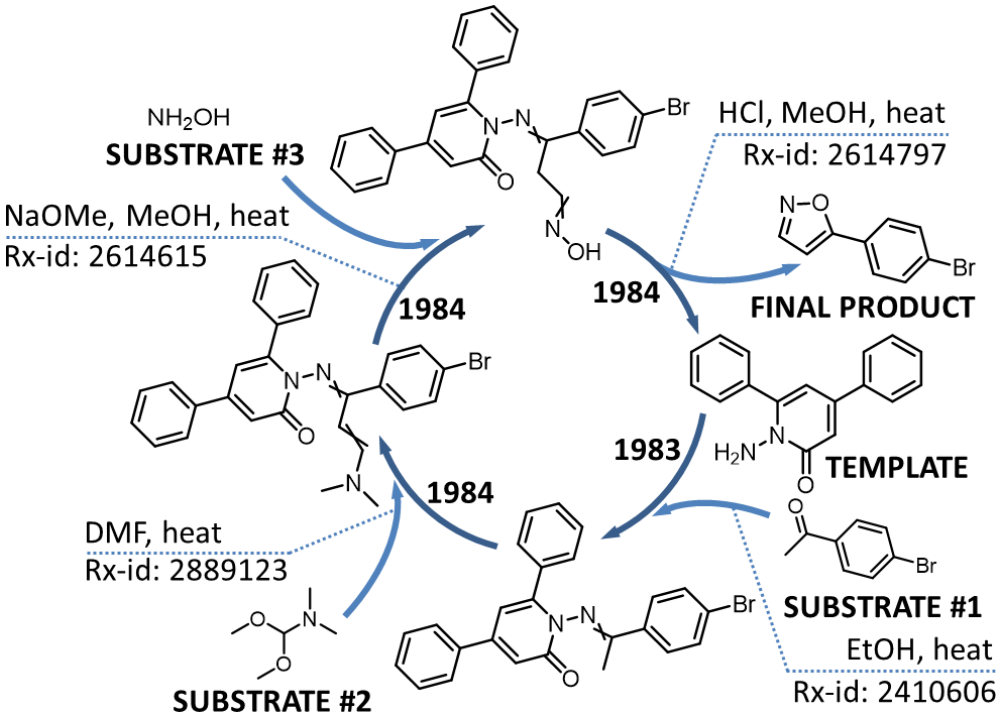


Figure 32. Another example of the Molina cycle. Structurally and functionally similar to the cycle from **Figure 29**, but producing 5-(4-Bromophenyl)isoxazole instead of 5-(4-chloro-phenyl)-1-phenyl-1*H*-pyrazole.

The initial intermediate of the last cycle (shown in **Figure 33**) has a two-fold role – during each completion, half of its molecules act as the template and the other half as the substrate. This effectively means that the "template" compound is never fully recovered unless this compound is replenished by external means, the cycle self-extinguishes. As seen in **Figure 33**, the cycle begins with esterification, followed by acyloin condensation – at this point two equivalents of the starting acid differentiate into the substrate and the template. The resulting benzoin acts as a benzoyl anion equivalent in α-acylation with trans-geranyl bromide. Following the reduction to diol, the periodate cleavage then leads to regeneration of one equivalent of benzoic acid, while the other equivalent remains incorporated into the newly formed product. The original literature entry for acyloin condensation (black rx-id) proposed by Cyclorg included the conditions that led to the mixture of products, thus more contemporary conditions have been proposed (red rx-id).

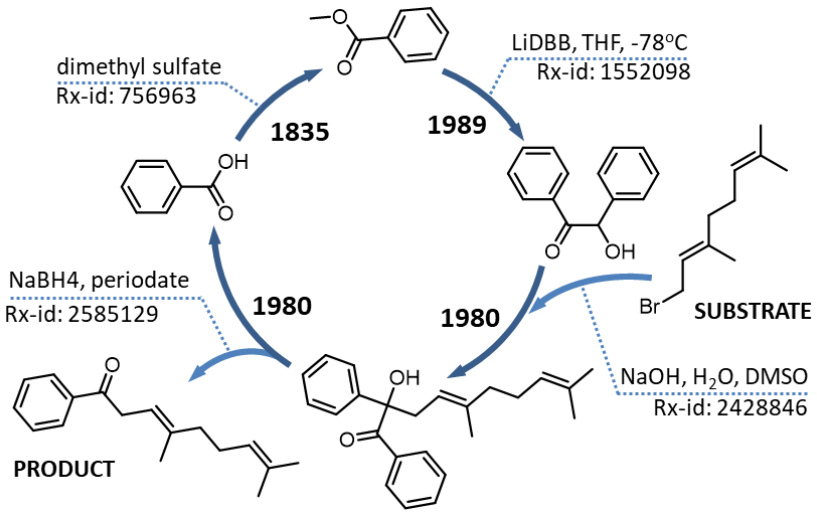


Figure 33. The cycle that took 154 years to complete. Benzoic acid acting both as the substrate and as the template compound to produce (E)-4,8-dimethyl-1-phenylnona-3,7-dien-1-one.

4.4.4. Cycle families

In addition to full cycles, Cyclorg also displays “inner shortcuts” (i.e., inner arrows interconnecting the vertices within a cycle). Those describe additional transformations possible between the intermediates involved. In fact, it is entirely possible to find a cycle within a cycle. This is best illustrated in **Figure 34**, where multiple sub-cycles are deduced from the graphical output of Cyclorg. This “family” of cycles illustrates the ability to recover and reuse useful reagents. The final product of all three is ethyl a-fluorocyclopentylidene acetate, which is a relatively expensive building block of drug candidates for type 2 diabetes.[91] This product is formed in a 59% yield by alkenylation of the 2-fluoro-3-hydroxy-2(organoselanyl)alkanoate precursor — at the same time, the diphenyl diselenide by-product is formed in a higher, 70% yield, which motivates its recovery into the cycle precursors. Among the “recovery routes”, the three-step sequence (reactions marked 3-4-5) is the least appealing as it requires stannate to be synthesized in a separate step[92] before reacting it with the α-bromo ester. The two-step sequence (reactions 6-5) generates the selenide anion *in situ* to condense with halide immediately, and does not require the use of the much more expensive ethyl chlorofluoroacetate reagent present in the one-step (reaction 7) recovery route.

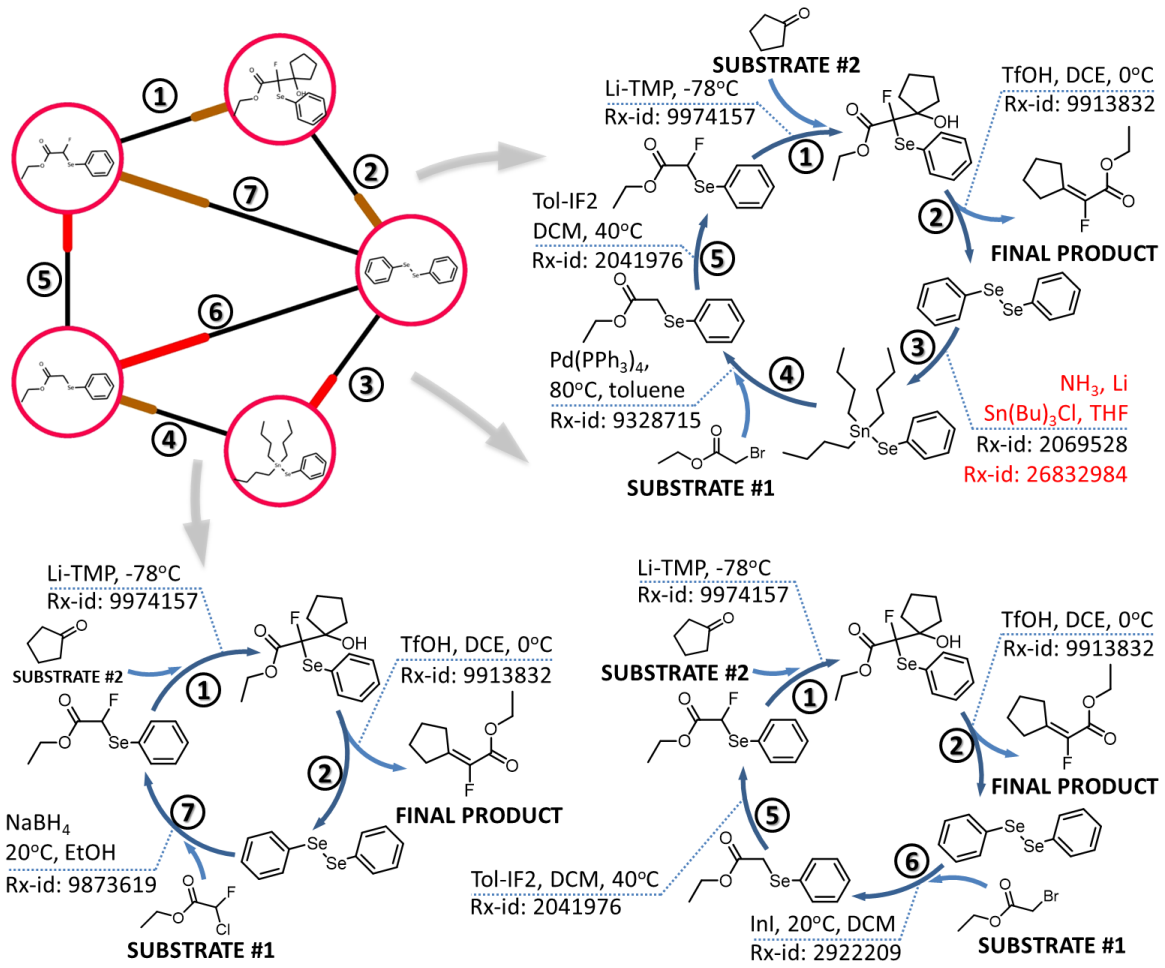


Figure 34. A cycle producing pharmaceutically relevant ethyl α-fluorocyclopentylidene acetate and recovering the phenylselenyl moiety via three different routes. The scheme on the left is a raw output from Cyclorg. The colored endings of the connections specify a 30-year period in which particular reactions were published.

Yet another example of a cycle family – shown in **Figure 35** – demonstrates how two similar cycles can produce drastically different products. Here, the additional cycle is deduced from one of the transformations being two-directional. Both cycles start from the same addition of triphenylphosphine to imidoyl chloride resulting in arylazomethylenetriphenylphosphorane, but each carries on differently to synthesize distinct products. In case of the three-step cycle, the resulting phosphorane undergoes a further reaction with N-methylhydroxylamine, followed by cyclization to give methyl 5H-imidazo[1,2-b]indazole-3-carboxylate. On the other hand, the two-step cycle performs the same cyclization one step earlier to give methyl 3,4-dihydro-4-oxoquinazoline-2-carboxylate. The latter is an example of a useful product made from the two-step cycle, as the synthesized 4-quinazolinone scaffold is present in

numerous drugs, such as Afloqualone, Albaconazole, Balaglitazone or Raltitrexed.[93,94] Triphenylphosphine, a compound which participates in both cycles, is a frequent template found in Cyclorg. One of the following sub-sections further elaborates on the triphenylphosphine-based cycles.

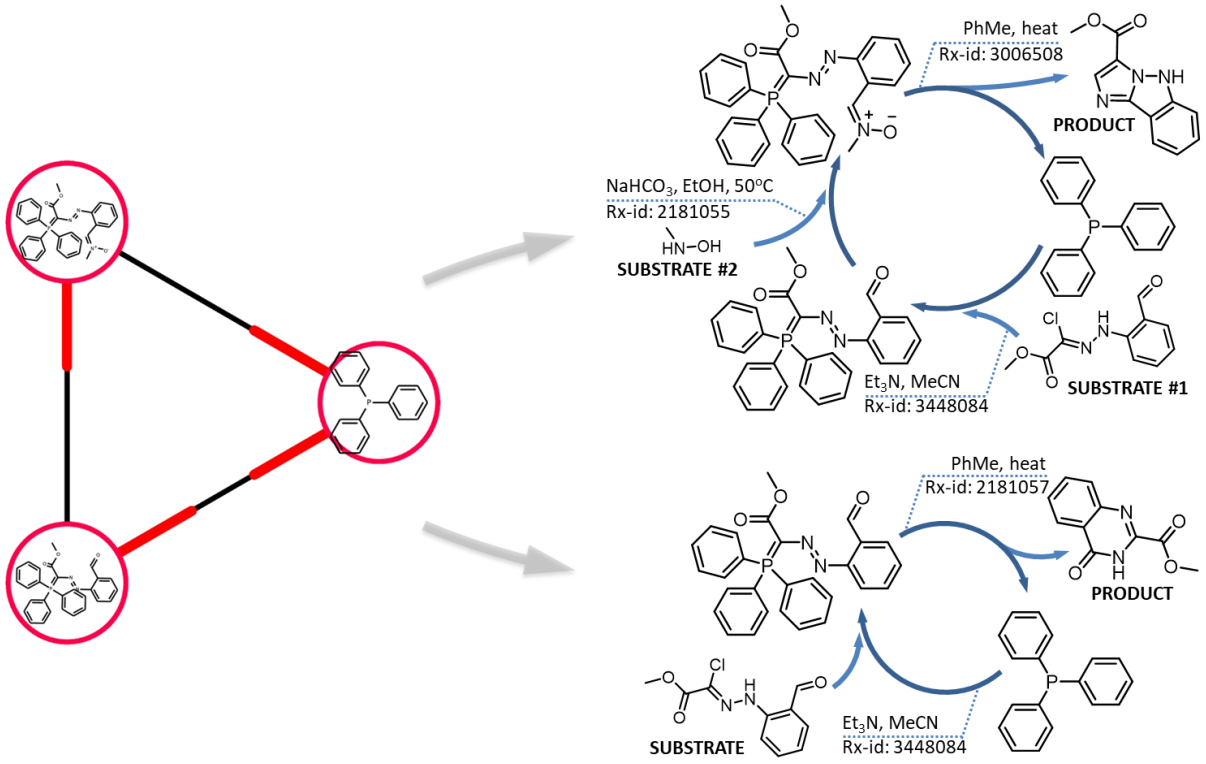


Figure 35. Two cycles comprising a family. The scheme on the left is a raw a output from Cyclorg. The colored endings of the connections specify a 30-year period in which particular reactions were published (see the color legend). If two ends are colored, it means that the reactions in both directions are known.

4.4.5 Chiral auxiliary-based cycles

The cycles based on asymmetric syntheses were, predictably, the most frequent amongst useful candidates, as many chiral auxiliaries are, by design, recoverable at the end. Technically, they are template-based cycles with said auxiliaries acting as templates – however, identifying such cycles is a mere formality and in most cases they comprise works of just one group – therefore, relatively few examples are presented in this thesis.

The most elegant example is shown in **Figure 36** and it is based on the works of the Senanayake group,[95] which - despite of its reliance on chiral auxiliaries – is not as apparent as other asymmetric synthesis cycles.

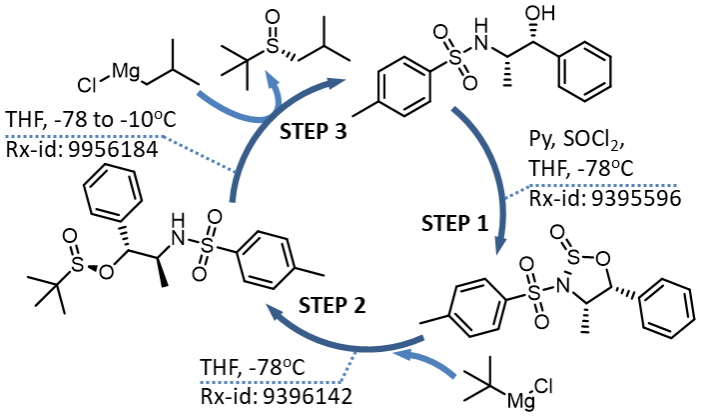


Figure 36. A three-membered cycle that inserts the sulfinyl group between two alkyl moieties with high stereoselectivity. N-((1R,2S)-hydroxy-1-phenylpropan-2-yl)-4-methylbenzenesulfoamide is used as a chiral agent – upon undergoing cyclization with $SOCl_2$, Grignard reagents are added sequentially to give chiral sulfoxide with high steroselectivity and the regenerated chiral auxiliary.

The cycle shown in **Figure 37** resembles the cycles based on intramolecular reactions. There, it produces (S)-methyl 2-formyl-6,6-dimethyl-2-cyclohexene-1-acetate from 2-formyl-4,4-dimethylcyclohex-2-en-1-one and bromoacetaldehyde diethyl acetal, using (S,S)-(+)-1,2-bis(2-methylphenyl)ethane-1,2-diol as the circulating chiral auxiliary. The cycle begins with the reaction of bromoacetal and 1,2-diol to give trans-2-bromomethyl-4,5-bis(2-methylphenyl)-1,3-dioxolane, followed by elimination of the bromide group. The resulting (S,S)-4,5-bis(2-methylphenyl)-2-methylene-1,3-dioxolane undergoes cycloaddition with 2-formyl-4,4-dimethylcyclohex-2-en-1-one (substrate #2) to give an epimeric dioxolanone. In the presence of L-selectride, it undergoes the elimination of hydroxyl, followed by hydrolysis of the resulting oxonium species – at this stage, the product is formed which, during the next step, is cleaved off and the 1,2-diol chiral auxiliary is regenerated.

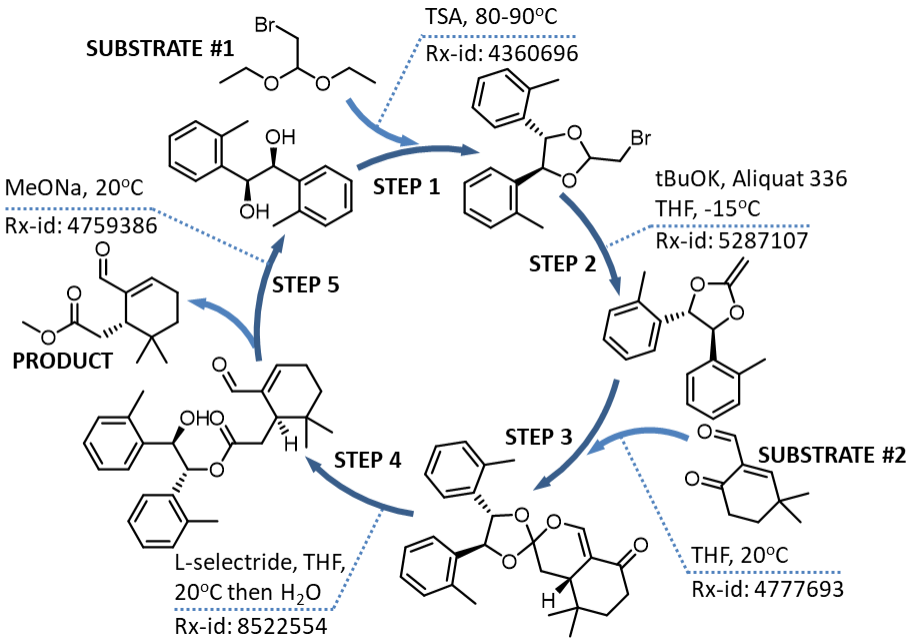


Figure 37. An example of a cycle based on the chiral auxiliary. The general method of using bromoacetals was published in 1995 by the Wallace group, while specific examples leading to this particular product were described later, in 1997 and 1999. In 2003, improved conditions were proposed for step two.

The example shown in **Figure 38** is similar to template-based cycles, where a product is assembled progressively on the template (in this case – the chiral auxiliary). Albeit technically developed by one group, the regeneration of template was specifically reported by another team. The cycle produces (S,Z)-3-hydroxy-2,2,4-trimethylhept-4-en-6-ynoic acid from 2-bromoisobutyric acid bromide and (Z)-2-methyl-5-(trimethylsilyl)pent-2-en-4-ynal. The Evans reagent, (4S)-4-isopropyl-1,3-oxalidin-2-one, is used as the chiral auxiliary that "circulates" through the cycle, regenerating after each completion. The cycle begins with alkylation of a chiral oxazolidinone, published by Kende in 1989. The next step involves an asymmetric aldol reaction, followed by rearrangement published by the same author a year later. The cycle ends with desilylation and hydrolysis leading to the formation of the product and – according to the work from 1990 – recovery of Evans oxazolidinone.

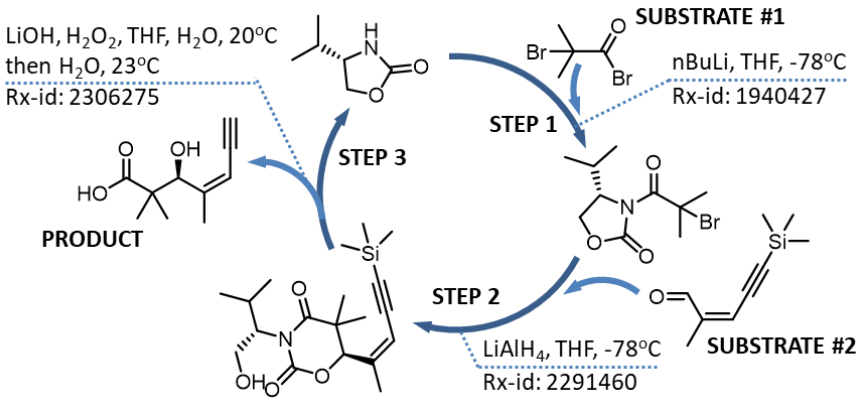


Figure 38. A three-membered cycle based on the Evans chiral auxiliary (4S)-4-isopropyl-1,3-oxalidin-2-one.

The cycle presented in **Figure 39** occupies the middle ground between intramolecular- and template-based cycles, i.e., while the product is being assembled progressively, this is interrupted by a group interconversion at steps 2 and 3. It produces (S)-1-((1S,3aS,6aS)-1,3a,4,6a-tetrahydropentalen-1-yl)ethane-1,2,-diol from bromoacetyl bromide, silver nitrate and (3aS, 6aS)-1,3a,4,6a-tetrahydropentalene. Bromoacetyl bromide is first substituted with an auxiliary, then with a nitrate. The latter works as an oxygen donor, as in the next step nitrate ester cleavage results in the formation of an aldehyde group. This group is attacked from the *re* side by tetrahydropentalene and the product is reductively cleaved off the auxiliary.

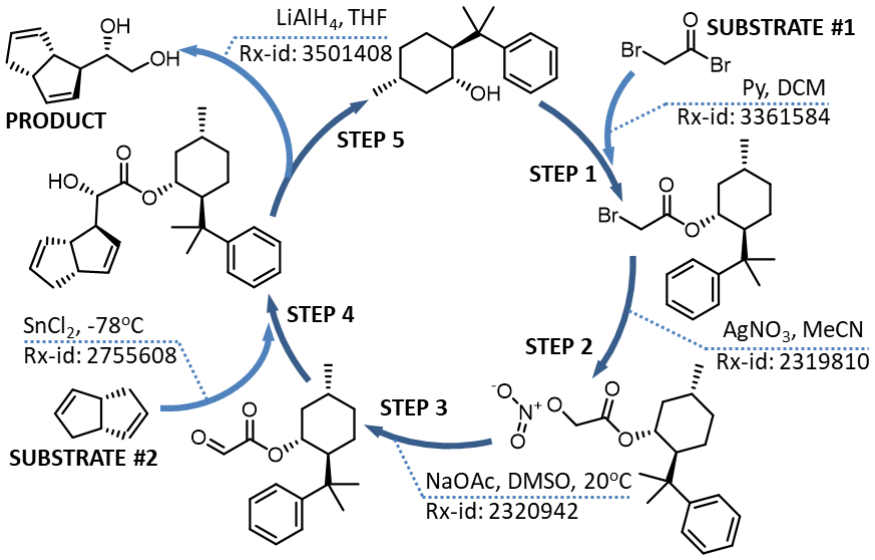


Figure 39. A five-membered cycle based on the (-)-8-phenylmenthol chiral auxiliary.

The cycle shown in **Figure 40** produces (S)-(+)-2,4-diphenylbut-3-yne-1,2-diol from 2,2-diethoxyacetophenone and phenylethynylmagnesium bromide. Cyclorg originally proposed conditions that lead to the mixture of the above-mentioned product and (E)-2,4-diphenylbut-3-ene-1,2-diol (not shown), thus alternative conditions have been proposed in their place.

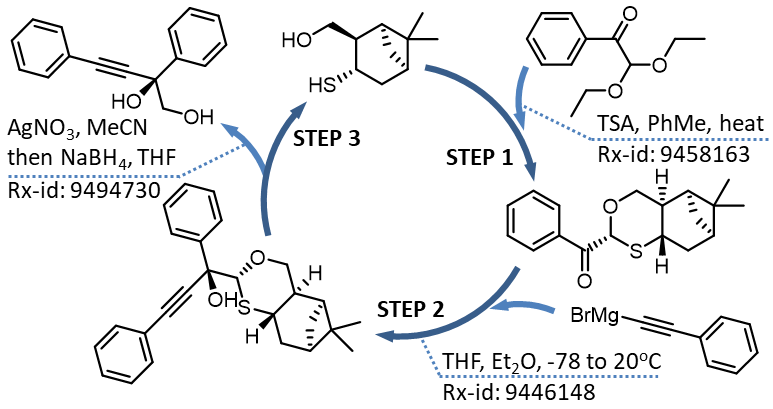


Figure 40. A three-membered cycle based on the 10-hydroxyisocamphenylthiol chiral auxiliary.

The example shown in **Figure 41** illustrates a cycle in which every intermediate is larger than the product. It produces methyl (S)-6-butyl-2-hydroxycyclohex-1-ene-1-carboxylate from ethyl 2-hydroxycyclohex-1-ene-1-carboxylate and n-butyllithium. (1R,2S,3R)-3-[N-Phenylsulfonyl-N-(3,5-dimethylphenyl)amino]bornan-2-ol circulates through the cycle as the chiral auxiliary, which is regenerated after each completion of the cycle.

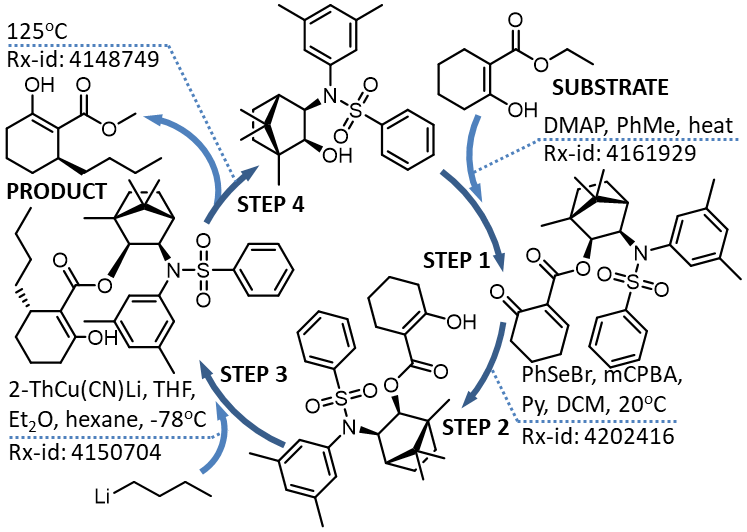


Figure 41. A four-membered cycle based on the (1R,2S,3R)-3-[N-Phenylsulfonyl-N-(3,5-dimethylphenyl)amino]bornan-2-ol chiral auxiliary.

4.4.6. Triphenylphosphine-based build-up cycles

Build-up cycles were frequently based on the triphenylphosphine template – as such, they typically synthesized large products and were often assembled from the works of numerous groups. Therefore, they are described here as a sub-category. Two useful examples have already been described in **Figure 35**. Another example, shown in **Figure 42**, took exactly one hundred years to complete. It produces a linear polyyne chain present in many natural compounds, including ichthyothereol but uses drastic reaction conditions of flash vacuum pyrolysis.

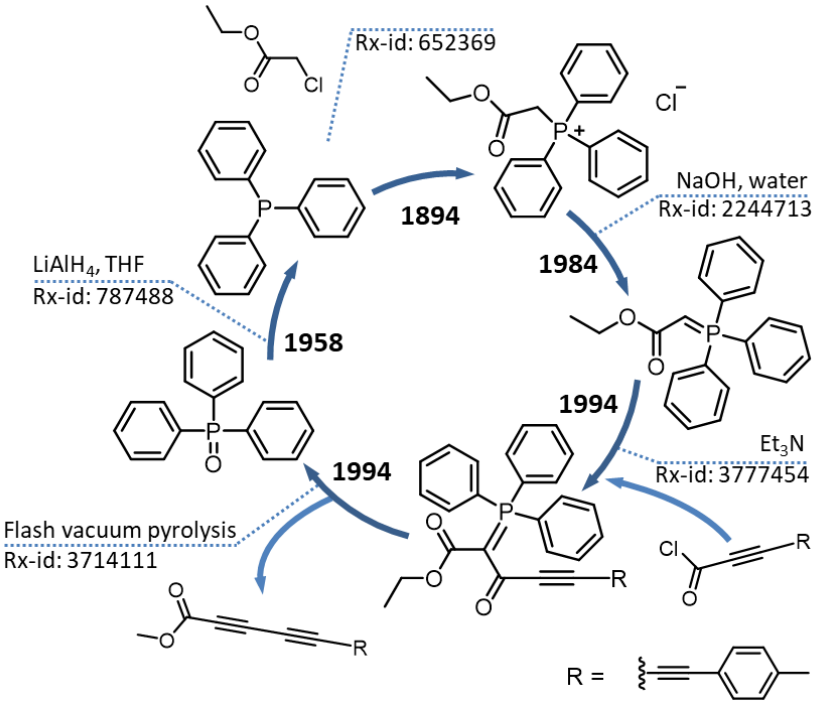


Figure 42. The cycle that took 100 years to complete. This cycle was initially opened in 1894 with the report of the synthesis of (ethoxycarbonylmethyl)triphenylphosphonium chloride and closed in 1994 with the publication describing the synthesis of linear polyynes from phosphonium ylides, indicated by the two red arrows at the bottom of the left section and the blue arrows in the middle section.

The next example (**Figure 43**) relies on milder conditions than pyrolysis, although all its steps were developed by the same group. It produces 3-cyclohexylideneisobenzofuran from 3-bromophtalide and cyclohexanone.

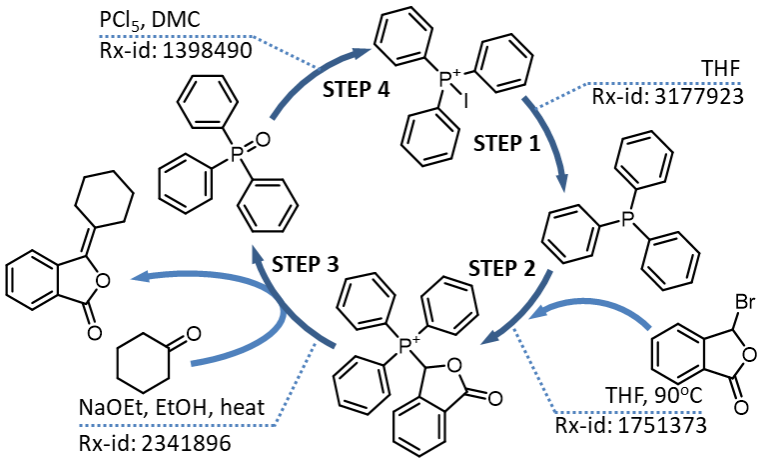


Figure 43. A four-membered cycle based on triphenylphosphine, all published in one article.

4.4.7. Biomimetic cycles

The wide interest in the chemical models of metabolic cycles was already noted in section **3.2**. The interest in the ornithine cycle, particularly, is reflected by the attempt made in 1990 to recreate all of its individual steps (shown in **Figure 1**) with the use of organic-chemical reactions.[96] In order to address the non-trivial issue of non-enzymatic transformation of ureido into guanidinium groups, the final model included an additional step and two protective groups. While this a deliberate effort to recreate the entire cycle, there are also other independent reports covering individual steps – these reports were successfully assembled into new cycle candidates and can be readily accessed from Cyclorg.

Figure 44 shows an interesting cycle comprising two metabolic cycles and a wide assortment of available reaction conditions. The ornithine cycle was recreated *in vitro* biochemically as well as chemically and, according to the rx-ids and/or other literature sources provided, it is possible to use different forms of prefabricated nitrogen as the main substrate of the cycle (e.g., O-methylisourea, cyanogen bromide or potassium cyanate). The nitric oxide synthesis cycle, which involves the same intermediates as the urea cycle, but in reverse order, produces nitric oxide essential to the functioning of advanced organisms.[97]

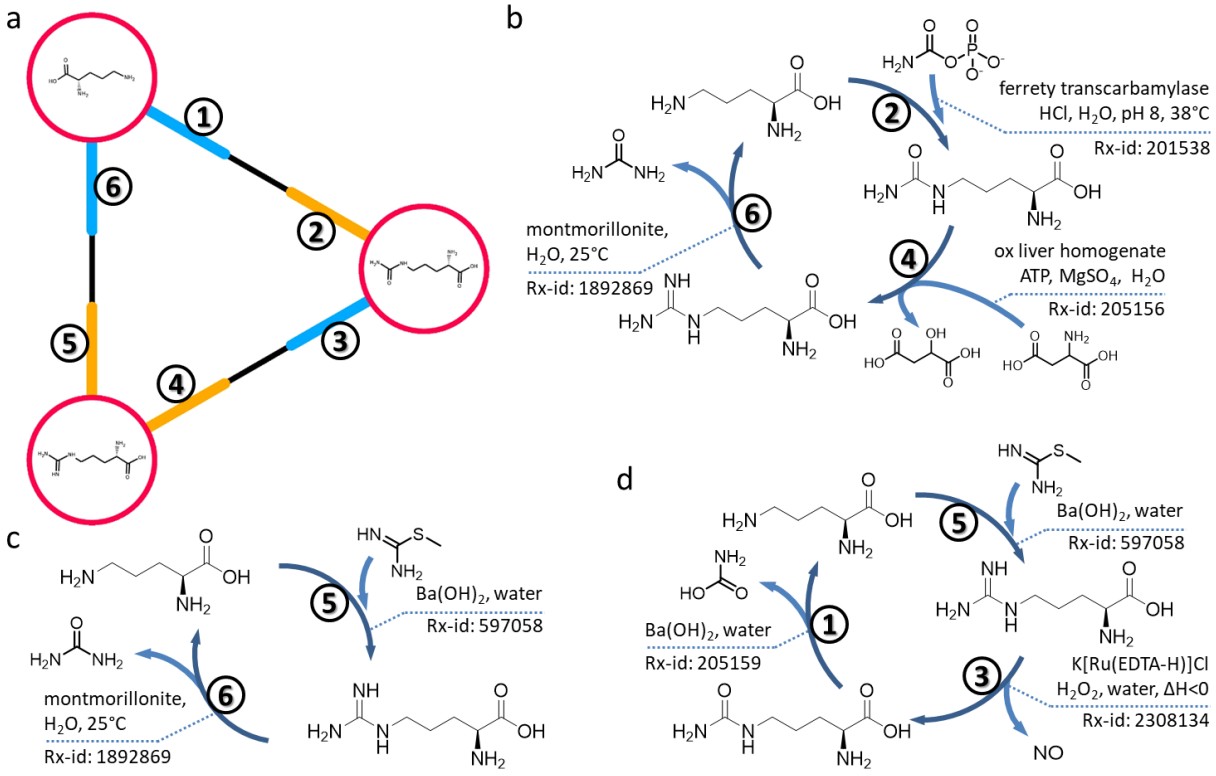


Figure 44. The artificial ornithine and nitric oxide synthesis cycles identified by Cyclorg. The raw output is shown in (a), with numbers assigned to each transformation. The *in vitro* recreation of the urea cycle, shown in (b), comprises both chemical (6) and biochemical (2 and 4) reactions, while the purely chemical-organic model, shown in (c), comprises reactions 5 and 6, whilst using methylisothiourea as an alternative source of nitrogen. The chemical model of the nitric oxide cycle is shown in (d).

Living cells are non-equilibrium systems, constantly exchanging matter with their surroundings. In this spirit, the biochemical methionine salvage cycle, shown in **Figure 3**, is relevant as only one thiomethyl group remains after every completion, while all the remaining matter is exchanged with the surrounding systems. The cycle shown in **Figure 45** is analogous to the one from **Figure 3** in the context that it "salvages" atoms from the used "cofactors" by incorporating them into newly assembled molecules of the substrate. The cycle starts with Pd/Zn-catalyzed coupling of benzyl bromide with benzoyl chloride. The resulting ketone undergoes condensation with p-methylbenzaldehyde, giving a pentaarylpentan-1,5-dione, followed by cyclization to a pyrylium bromide. Subsequently, the product from the previous step reacts with benzylamine yielding 4-(p-tolyl)-1-benzyl-2,3,5,6-tetraphenyl pyridinium bromide, which is subjected to pyrolysis resulting in the formation of benzyl bromide (the substrate used in the first step) and 4-(p-tolyl)2,3,5,6-tetraphenylpyridine.

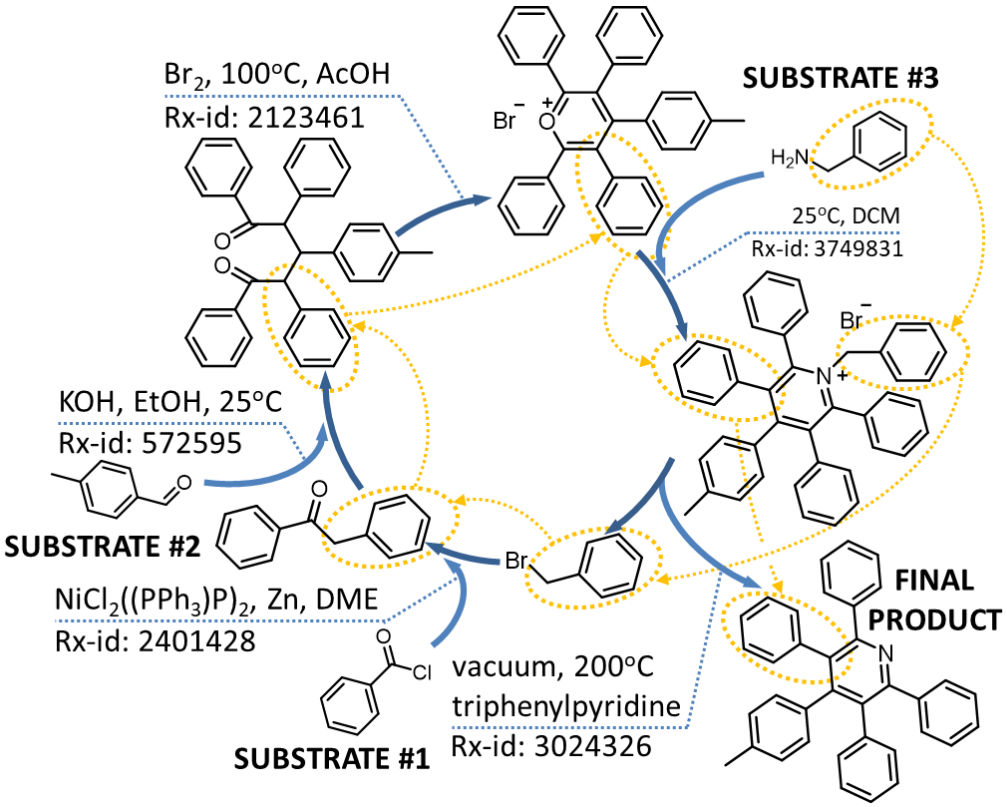


Figure 45. In this example, not a single atom remains in the cycle for longer than one full completion – all matter that enters leaves. The blue arrows indicate reactions, whereas the orange arrows and circles are used to highlight the movement of the benzyl group – first introduced as a reactant, then incorporated into the main scaffold, and ultimately leaving in the cycle product.

Figure 46 shows a false-positive candidate. Although very similar to the one illustrated in **Figure 45**, this cycle is flawed by a certain inconsistency with regard to the counter ions on the cationic intermediates – pyrylium is first produced as a salt of perchloric acid, but in the subsequent step it is supposed to be used as a salt of bromine. Cyclorg recognizes salts as two separate species – if the structure of the main reagent is a match, the counter-ions are neglected. This leaves many cycles in need of further fine-tuning – while including additional steps solves this particular problem (allowing me to present this cycle), this would be done outside of Cyclorg expertise. If future cycles are to be discovered with literature-navigating algorithms, such issues should be taken into account and software should be improved to recognize reagents (salts) more holistically, and not just as individual molecules or counter-ions without a pair.

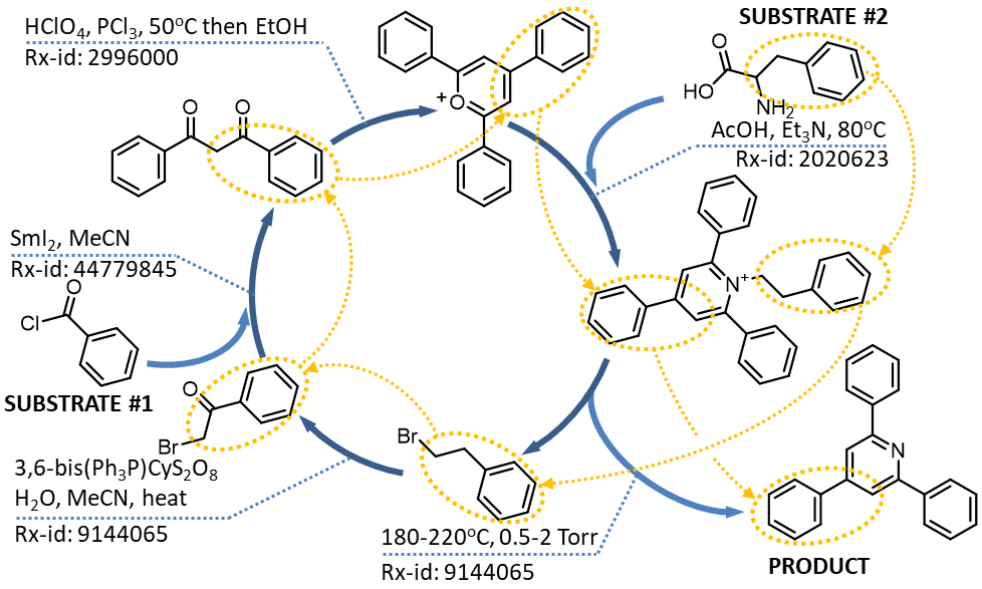


Figure 46. A false-positive example of the cycle, where all the atoms leave after one full completion. Pyrylium is produced as perchlorate, but subsequently it is used as bromide in the following step. The blue arrows indicate reactions, whereas the orange arrows and circles are used to highlight the movement of the benzyl group.

4.4.8. Moebius-like cycle

The potential of Cyclorg as an "idea generator" is well-illustrated by the discovery of a cycle that was originally inspired by a trivial error in the underlying NOC database.

When the cycle candidate shown in **Figure 47a,b** was first displayed on results page, I recognized a discrete symmetry in some of its intermediates – the structure comprising two phenyl glycine amino acid derivatives joined together by the mutual amino group created a prospect of indirect auto-amplification (i.e., compound promotes synthesis of its close precursor). Unfortunately, upon inspection, it turned out to be a linear pathway erroneously identified as cycle. The author(s) of the relevant NOC database entries forgot – or neglected – to specify stereochemistry on phenylglycinol. Because of such a trivial error, the 5-step discontinuous pathway starting from S-phenylglycinol and leading to R-phenylglycinol was closed into a cycle with racemic phenylglycinol as the intermediate – see **Figure 47b,c**.

Before the "cycle" was discarded, I realized that it was entirely stereospecific – i.e., if only the configuration of the starting substrate (phenylglycinol) were changed, the configuration of the product would also change. Therefore, starting from S stereoisomer would yield R stereoisomer and starting from R-substrate would result in S-product – it is a total inversion, which could also work for mixtures of isomers. By extension, completing this pathway twice returns the substrate to its initial configuration – so it was, in fact, a symmetric cycle where the configurations interchange at the "twist" slightly reminiscent of the famous Moebius strip (as illustrated in **Figure 47d**) – hence the nickname of this cycle.

In its current form, the cycle was practically futile as the consumption of benzaldehyde results in its recreation after three steps. However, this is not the same benzyl group as there was an actual exchange of two identical groups. This means a possibility of interconversion – so if, for example, the substrates were phenylglycinol and some Ar-substituted benzaldehyde, then an Ar-substituted glycinol and unsubstituted benzaldehyde would be produced three steps later. This potential for interconversion is illustrated in **Figure 47e** by the R_1 and R_2 Markush structures. In effect, not only can this symmetric cycle invert configuration of amino acids, but also alter their substitution pattern. An experimental validation is, of course, required, but this cycle could potentially be used as a flexible method for the synthesis of modified amino acids.

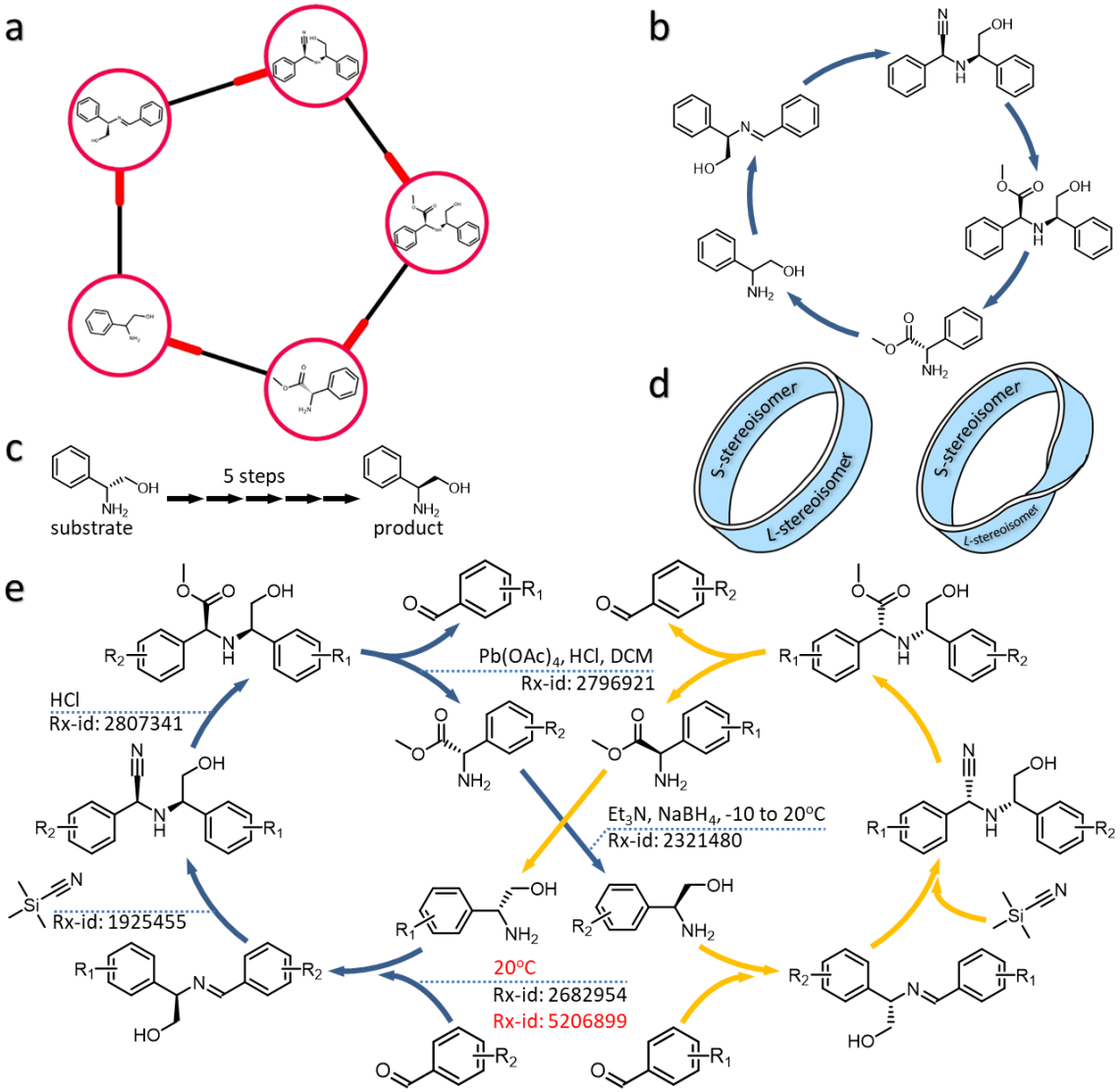


Figure 47. The "Moebius" cycle was designed from the Cyclorg entry shown in (a). Initially identified as a cycle presented in (b), it was actually a linear pathway transforming chiral S-phenylglycinol into its opposite enantiomer (c). The simplistic overview of this inversion as well as the inspiration for the nickname of the cycle is shown in (d), where the stereospecific cycle is shown on the left and the "Moebius" cycle is shown on the right. The analysis of this candidate inspired the design presented in (e). Its application is two-fold. First, it changes the substitution patterns of the compounds introduced into the cycle (keep track of the R_1 and R_2 Markush) – which allows for creating entire libraries of chemical analogues. Second, it reverses the configuration of the amino acid derivatives introduced. All the reactions in the cycle are stereoselective – no stereocenters are introduced from outside the cycle. All the reactions can be performed in MeOH or its mixtures, red rx-id provides updated conditions compatible with this solvent. The arrows intersecting in the middle show that both ,,blue" and ,,orange" transformations could work one-pot, independently of each other (if given enough $NaBH_4$).

4.4.9. Self-extinguishing cycles

Some cycles are characterized by their self-extinguishing activity – i.e., if such a cycle is not constantly fed with one of its "core" intermediates (in addition to all substrates), the total concentration of all its intermediates will diminish after each completion until the cycle ceases operation. One such example has been shown in **Figure 33**.

Another example is shown in **Figure 48**, where one of the intermediates – pyridine-2-carbaldehyde – works as both the substrate and the template. Thus, every completion results in its disproportionation – one equivalent leaves with the product, while the other is recovered for another run. Here, two molecules of pyridine-2-carbaldehyde react with tetrahydroisoquinoline to generate racemic oxazolidine, which further reacts with N-methylmaleimide to produce a racemate of a polycyclic four-stereocenter endo-cycloadduct (endo:exo 2.1:1), whose central ring system is known to exhibit antidepressant activity.[98]

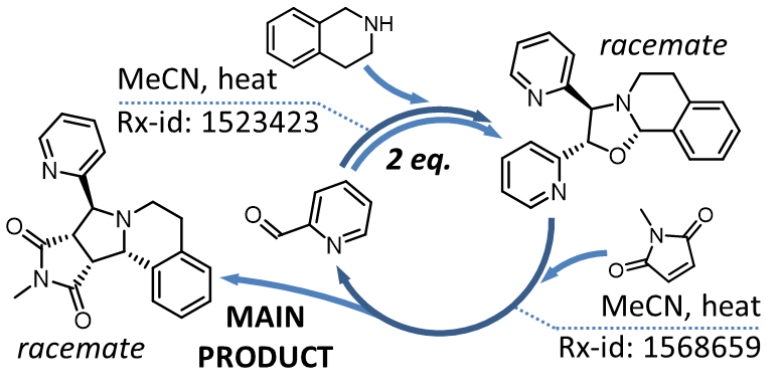


Figure 48. A self-extinguishing cycle. In this cycle, one pyridine-2-carbaldehyde is regenerated — since two equivalents were used during the synthesis, the cycle will self-extinguish if the substrate is not continuously fed. Both steps were performed in acetonitrile making this cycle a viable candidate for one-pot operation.

4.4.10. Miscellaneous short small-molecule cycles

Some cycles do not belong to any of the presented categories. **Figure 49** presents a cycle which was rather unintentionally closed by the Rafiqul group and published completely in one article.[99] Here, pivalaldehyde – acting as a template – reacts with substrate thioamide. Then, the resulting 2,6-di-tert-butyl-4-phenyl-6H-1,3,5-oxathiazine undergoes oxidation of the sulfide group. It is the spontaneous decomposition of the last intermediate that closes the cycle, resulting in formation of its by-product and regeneration of the "template" aldehyde. Decomposition proceeds through 3,5-diphenyl-5H-1,2,4-oxathiazole (not shown), which can be retrieved before it further transforms.

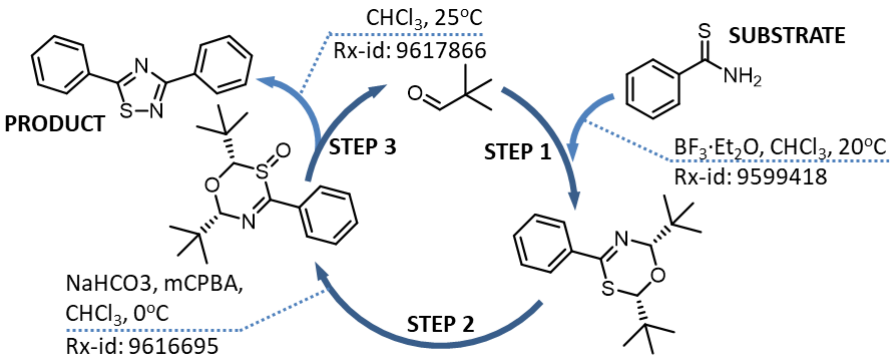


Figure 49. A three-membered cycle that produces 3,5-diphenyl-[1,2,4]thiadiazole benzenecarbothioamide through slow decomposition of intermediate 2,6-di-tert-butyl-4-phenyl-6H-1,3,5-oxathiazine S-oxide. Perhaps being an unintended effect, the product of step 2 decomposes slowly at room temperature to give both product and starting pivalaldehyde.

The cycle in **Figure 50** was originally identified by Cyclorg as a 4-membered cycle (the lacking step is 2,4-dinitrophenolate anion), however it could be interpreted as 2-membered cycle from chemical point of view.

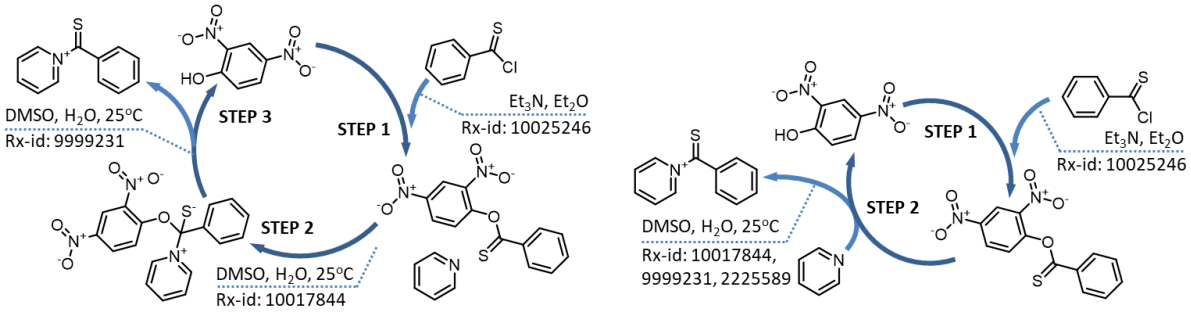


Figure 50. An example of the template cycle that produces 1-(phenylcarbothioul)pyridine-1-ium from thiobenzoyl chloride and pyridine by building it on 2,4-dinitrophenol "template". A strongly electron-withdrawing dinitrophenol moiety enables nucleophillic substitution of pyridine for the thioacyl group and acts as the leaving group, regenerating itself after every cycle completion.

4.4.11 Cycle regenerating two templates

The example shown in **Figure 51** is a rather flexible cycle, as it could work with some variety of substrates. More importantly, it reuses not one but two of its reagents – trimethylsilyl chloride and methyl propionate. The cycle is conceptually interesting even if its products could be obtained more efficiently by means of the contemporary alkylation methods.

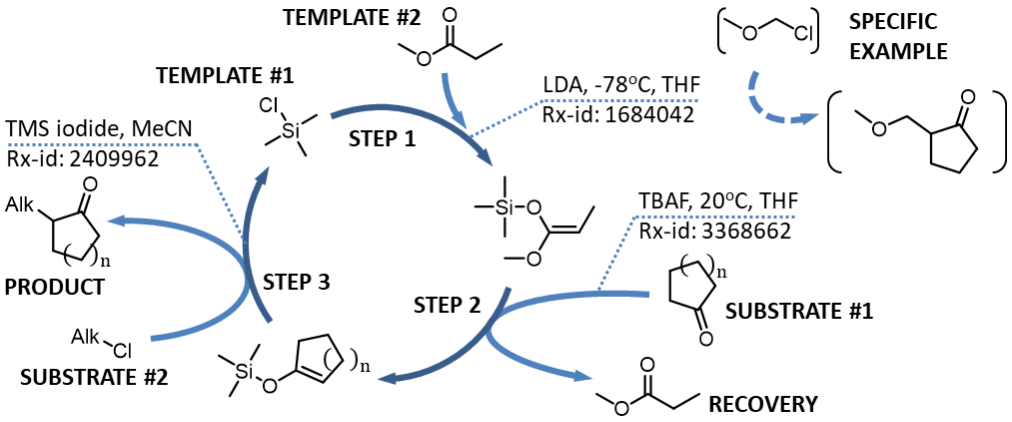


Figure 51. A cycle that uses and then recovers two different template compounds - trimethylsilyl chloride and methyl propionate. According to literature, *n* = 1,2.

4.4.12. Catabolic cycle

Not all biochemical cycles produce large compounds – some are intended as a method of disposal. The cycle shown in **Figure 52** holds some resemblance to the urea cycle presented in **Figure 1** in that it transforms acryl chloride into less reactive compounds. It is also notable for taking the most time of all the presented cycles to close – exactly 160 years.

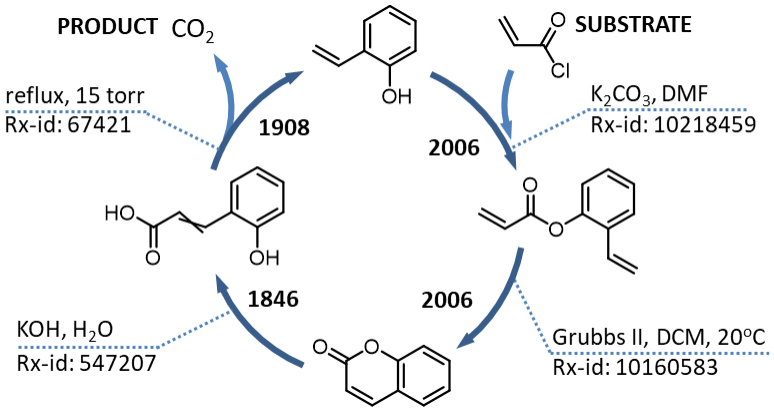


Figure 52. The cycle that took 160 years to complete. It decomposes an introduced compound into gaseous side-products – ethylene (not shown) and carbon dioxide. Such cycles are of little application to an organic chemist, but display a certain level of biomimicry.

5. Kinetix-based research

As previously described in section **3.5**, the systems giving rise to chemical oscillations are well-studied and have attracted growing interest. On the other hand, how systems quench or otherwise react to the oscillations introduced from outside is largely unexplored, yet relevant to a number of fields. For example, knowing the system capability for suppressing oscillations could shed light on homeostasis or help establish the maximum interval at which a chemical reactor could be safely loaded with substrates. Knowing what and when chemical networks are transmissive to time-varying signals could provide a valuable insight into a *de novo* design of chemical logic gates and biomimetic signaling pathways.

The initial choice of the subject has been inspired by one rather unexpected discovery – that whether chemical reactions can pass substrates' oscillations on to their products depends on frequency and that the boundary between frequencies of complete transfer and those of zero transfer is remarkably sharp.

The following parts of this section describe the initial experiment that led to this discovery (**5.1.**) and the following investigations (**5.2.** and **5.3.**). The explicit results are described in section **6**.

5.1. Introductory experiments

When I first gained access to Kinetix, I asked myself a question "What would happen with a chemical network if the substrate were fed in an oscillatory manner?" To answer this question, I designed a simple model system – a sequence of ten 1st order reactions, where the product of each is also the substrate in the following. The concentration of an initial reagent would oscillate and the last reaction would serve as an outflow:

$$a_0 \rightarrow a_1 \rightarrow a_2 \rightarrow a_3 \rightarrow a_4 \rightarrow a_5 \rightarrow a_6 \rightarrow a_7 \rightarrow a_8 \rightarrow a_9 \rightarrow a_{10},$$

$$\frac{d[a_n]}{dt} = k_n[a_{n-1}] - k_{n+1}[a_n], \qquad [a_0] = \sin \omega t + 2,$$

$$[a_{1-9}]_{init} = 1, \qquad [a_{10}] = const,$$

$$k_1 = 1, k_2 = 1, k_3 = 0.5, k_4 = 2, k_5 = 1, k_6 = 1, k_7 = 1, k_8 = 1, k_9 = 1, k_{10} = 0.5 \ \left[1/s\right].$$

The a_n denotes a participating reagent – the system is purely hypothetical, so at this point the participating species do not relate to any specific substances. The same goes for the reaction rates – in fact, at this stage the constants could be chosen rather haphazardly. Now, to keep this model simple, the concentration of the initial substrate – a_0 – is "fixed". This means that despite apparently being consumed in the first reaction, a_0 concentration (at the time *t*) is always equal to $\sin \omega t + 2$ and does not decrease from a_0 consumption – as if a_0 were fed from an infinitely large reservoir. It could be interpreted that reaction $a_0 \rightarrow a_1$ involves transport into a flow reactor of limited capacity or describes diffusion, where a_1 denotes a_0 transported into a new vessel.

Mathematically speaking:

$$\frac{-d[a_0]}{dt} = \omega \cos \omega t\,, \qquad \frac{-d[a_0]}{d[a_0]} = \frac{-d[a_0]}{d[a_1]} = 0.$$

This is one of the available approximation methods of Kinetix. The reactant concentrations that are not constant in time, but are still independent of the reactions that they are involved in are referred to as "fixed" in Kinetix – they always equal current values of the preassigned functions (of time *t*).

The result is seen in **Figure 53** – the concentration of the product oscillates with the same frequency as the substrate, albeit with a lower maximum amplitude. With time, the oscillations become regular – without a time-dependent input, the system would reach a stationary state at this point, which is now prevented by the external oscillations. Since the signal transferred from outside is the only concentration change observed in this state, I will refer to this state as "pseudo-stationary." In this state, all concentrations change periodically allowing for estimation of the average concentration values.

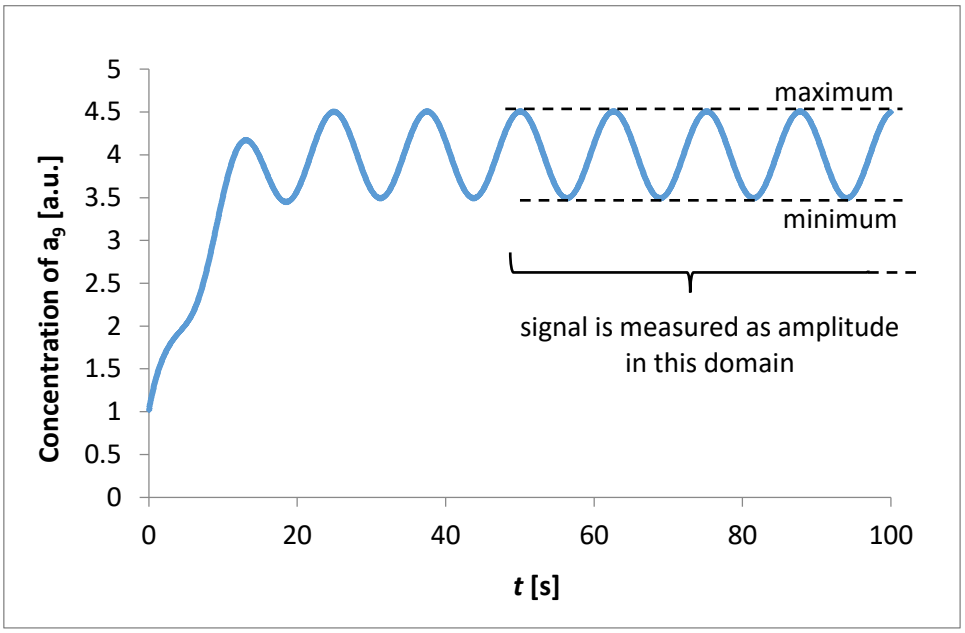


Figure 53. The concentration profile of the product of the ninth reaction in a sequence of ten 1st order reactions for ω=0.4 s^{-1}, f=0.064 s^{-1}. Amplitude is used as a measure of strength of the oscillatory signals and equals half of the difference between the maximum and minimum concentration, whilst oscillations are regular/periodic. The rate constants are $k_1 = 1, k_2 = 1, k_3 = 0.5, k_4 = 2, k_5 = 1, k_6 = 1, k_7 = 1, k_8 = 1, k_9 = 1, k_{10} = 0.5\ [s^{-1}]$.

Next, I extended the experiment to a range of frequencies, from $8 \cdot 10^{-4}$ to 1.6 Hz. As seen in **Figure 54,** at low frequencies (up to $5 \cdot 10^{-3}$ Hz) the maximum amplitude of product was almost equal to that of the substrate, while at high frequencies (more than $8 \cdot 10^{-2}$ Hz) it was near zero – the most conspicuous is the relatively sharp transition between these two states.

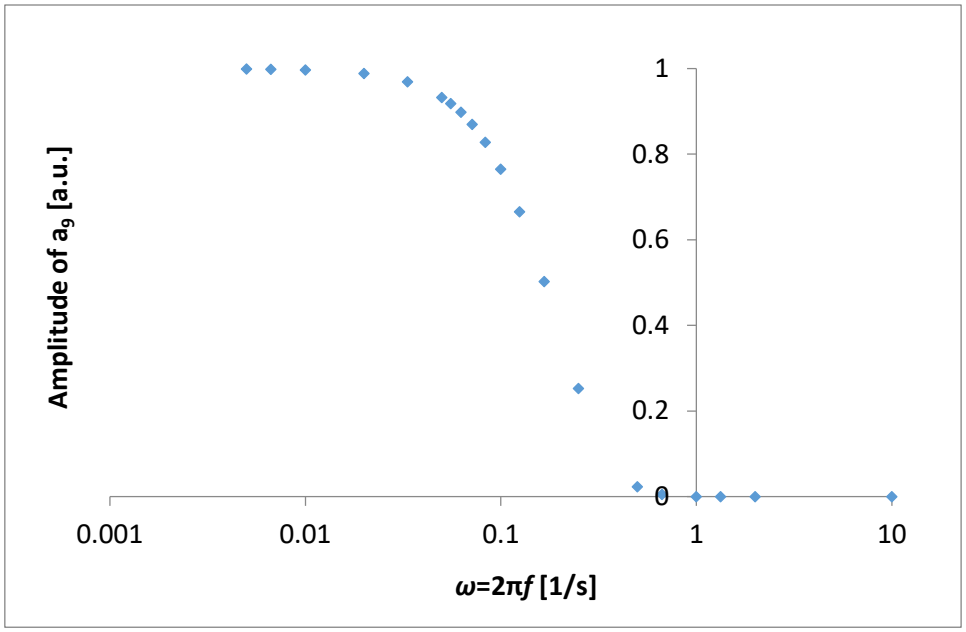


Figure 54. Strength of the propagated signal as a function of its frequency. Measured as the concentration amplitude of *a9* from the sequence of the 1st order reactions. $k_1 = 1$, $k_2 = 1$, $k_3 = 0.5$, $k_4 = 2$, $k_5 = 1$, $k_6 = 1$, $k_7 = 1$, $k_8 = 1, k_9 = 1$, $k_{10} = 0.5$ $[s^{-1}]$. The amplitude of a_0 is 1.

The simulated chemical network was either fully transmissive or not at all transmissive, with only a narrow range of frequencies at which the signal was partially transmitted. In order to see if this phenomenon persists in other networks, I repeated the simulations, but changed the rates of the reactions involved. At first, the rate constants were chosen randomly:

$$k_1 = 1, k_2 = 1, k_3 = 1, k_4 = 2, k_5 = 2, k_6 = 3, k_7 = 1.5, k_8 = 1, k_9 = 1, k_{10} = 0.5 \ \left[{}^{1}/_{s}\right]$$

for the second network, and for the third:

$$k_1 = 0.5, k_2 = 0.5, k_3 = 0.5, k_4 = 0.25, k_5 = 1, k_6 = 1,$$

$$k_7 = 1, k_8 = 0.5, k_9 = 0.5, k_{10} = 0.5 \left[{}^{1}/_{s}\right].$$

The results of both new models are shown in **Figure 55**.

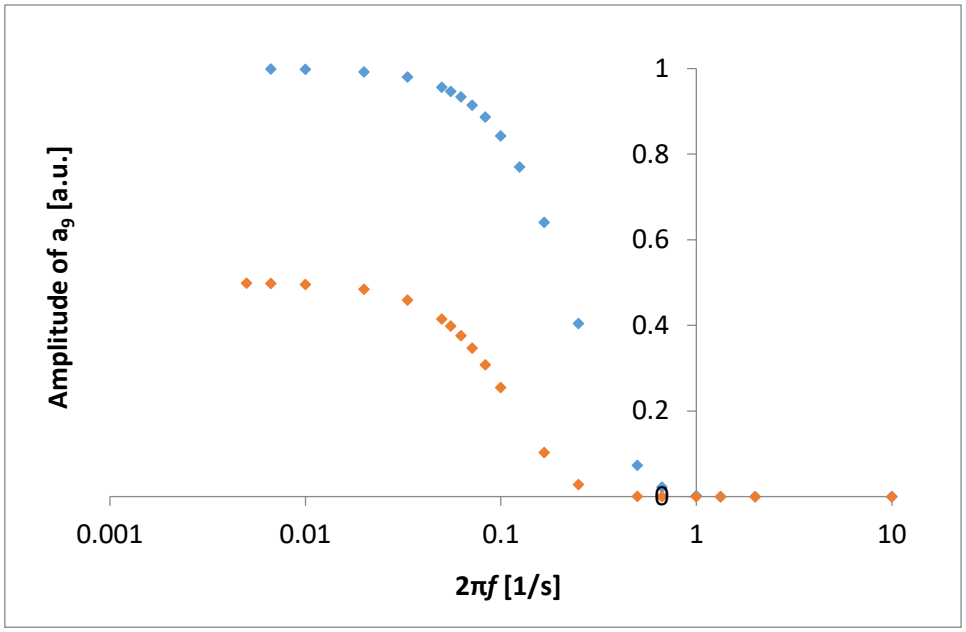


Figure 55. Strength of the propagated signal as a function of its frequency. Measured as the concentration amplitude of *a9* from the sequence of the 1st order reactions. The blue diamonds correspond to $k_1 = 1,\ k_2 = 1,\ k_3 = 1,\ k_4 = 2,\ k_5 = 2,\ k_6 = 3,\ k_7 = 1.5,\ k_8 = 1, k_9 = 1,\ k_{10} = 0.5\ [s^{-1}]$. The orange diamonds correspond to $k_1 = 0.5,\ k_2 = 0.5,\ k_3 = 0.5,\ k_4 = 0.25,\ k_5 = 1,\ k_6 = 1,\ k_7 = 1,\ k_8 = 0.5, k_9 = 0.5,\ k_{10} = 0.5\ [s^{-1}]$. The amplitude of a_0 is 1.

Only changing the reaction rates, but simultaneously retaining the linear architecture, resulted in a sharp transition similar to that from the first network – albeit at different frequencies and in case of the third model the signal was only half as strong. I suspected that this decrease in amplitude was most likely the result of a different ratio between specific rate constants. Thus, in order to exclude that the maximum amplitude was affected by the absolute values of k_n, I have multiplied each of the rate constants from the initial model (**Figure 54**) by three:

$$k_1 = 3, k_2 = 3, k_3 = 1.5, k_4 = 6, k_5 = 3, k_6 = 3, k_7 = 3, k_8 = 3, k_9 = 3, k_{10} = 1.5\ \left[{}^{1}/_{s}\right].$$

As seen in **Figure 56**, increasing the reaction rate evenly across the entire chain did not change the amplitude of the last step, which means that the strength of signal must depend (apart from the starting value) on some ratio between the involved reaction rate constants.

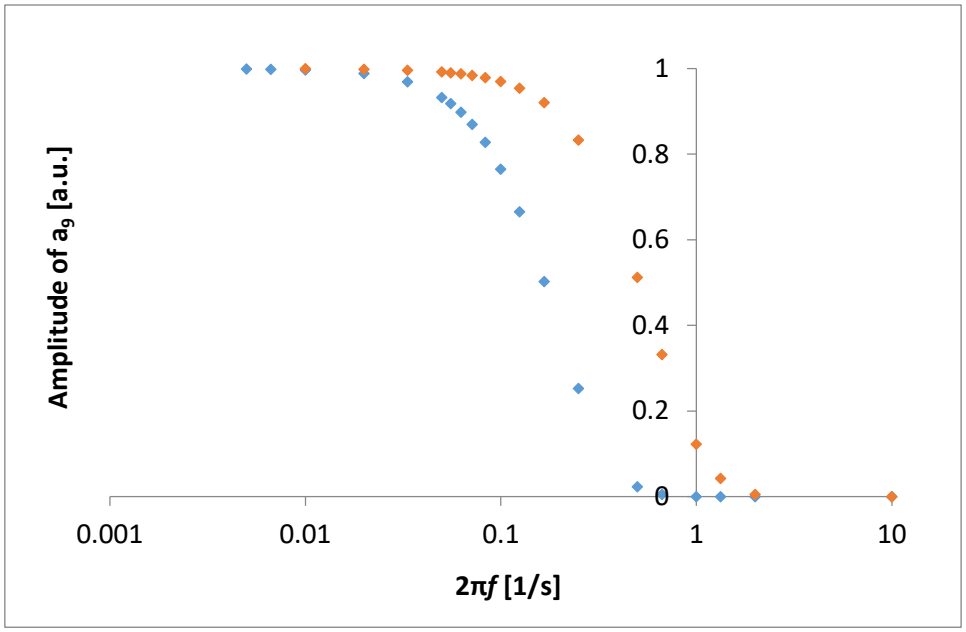


Figure 56. Strength of the propagated signal as a function of its frequency. Measured as the concentration amplitude of *a9* from the sequence of 1st order reactions. The blue diamonds correspond to the original from **Figure 54**: $k_1 = 1, k_2 = 1, k_3 = 0.5, k_4 = 2, k_5 = 1, k_6 = 1, k_7 = 1, k_8 = 1, k_9 = 1, k_{10} = 0.5\ [s^{-1}]$. The orange diamonds correspond to the "tripled" set: $k_1 = 3, \quad k_2 = 3, \quad k_3 = 1.5, \quad k_4 = 6, \quad k_5 = 3, \quad k_6 = 3, \quad k_7 = 3, \quad k_8 = 3, k_9 = 3, k_{10} = 1.5\ [s^{-1}]$. The amplitude of a_0 is 1.

At this point, I had a newly identified characteristic property of chemical systems, but no established rules governing it. In order to discover how the architecture and kinetic parameters of chemical networks relate to the maximum amplitude and the threshold frequencies, I began a systematic, bottom-up analysis. By observing the oscillations in various systems and situations – and under the helpful co-supervision from Prof. Robert Hołyst – I progressively gained understanding of their inner workings and one-by-one constructed the laws and equations that describe them.

5.2. Systematic analysis of linear reaction networks

According to the concepts of systems chemistry, I took a bottom-up approach to the subject – beginning from the simplest networks and linear kinetics, I worked my way up from hypothetical to realistic models.

This section describes step-by-step the process by which I came to an understanding of the signal propagation in the straight sequences of the 1st order reactions. First, I discovered how to predict the amplitude of the transferred signal in such systems (sub-section **5.2.1.**), then how to predict the threshold frequency at which the signal becomes dampened (**5.2.2.**). For explicitly presented results of the Kinetix-based research, see section **6.**

The subsequent chapters describe how I use this knowledge as the base for my analysis of more complex network elements. Section **5.3.** introduces the model of a simple enzymatic cascade, which is fully non-linear. In order to reliably solve such cascades, I first analyzed non-linear reaction kinetics in the well-understood linear reaction sequences (**5.4.1**), then I studied the cascade network architecture comprising only the well-understood 1st order reactions (**5.4.2**). Both non-linear kinetics and non-linear architecture, as well as some elements of linear sequences, are merged together in the realistic model of Ras/ERK cascade illustrated in section **5.5.**

All the described oscillations were measured after long simulation times, i.e., measured after the moment when the system typically reaches a stationary or equilibrium state. Since the introduced oscillations were sustained, they became periodic after the said moment (i.e., there were no other changes in the concentrations apart from the sustained oscillations). Such a state of periodic changes is referred to as “pseudo-stationary” in this work.

5.2.1. Predicting signal amplitude

The question that comes to one's mind when thinking about reaction rates and oscillations is "what would happen if there were a significant difference in the rates of two subsequent reactions." To answer this question, I created a chain of 30 reactions – with 15 slow reactions in a row, followed by 15 that were significantly faster:

$$a_0 \xrightarrow{slow} a_1 \xrightarrow{slow} a_2 \xrightarrow{slow} \dots \xrightarrow{slow} a_{13} \xrightarrow{slow} a_{14} \xrightarrow{slow} a_{15} \xrightarrow{fast} a_{16} \xrightarrow{fast} \dots \xrightarrow{fast} a_{28} \xrightarrow{fast} a_{29} \xrightarrow{fast} a_{30}$$

$$\frac{d[a_n]}{dt} = k_n[a_{n-1}], \qquad [a_0] = \sin 0.4t + 2, \qquad [a_{1-29}]_{init} = 1, \qquad [a_{30}] = const,$$

$$k_{1-15} = 0.5, \quad k_{16-30} = 2 \quad [1/s].$$

As seen in **Figures 57-58**, the signal was almost completely dampened at the junction between "slow" and "fast".

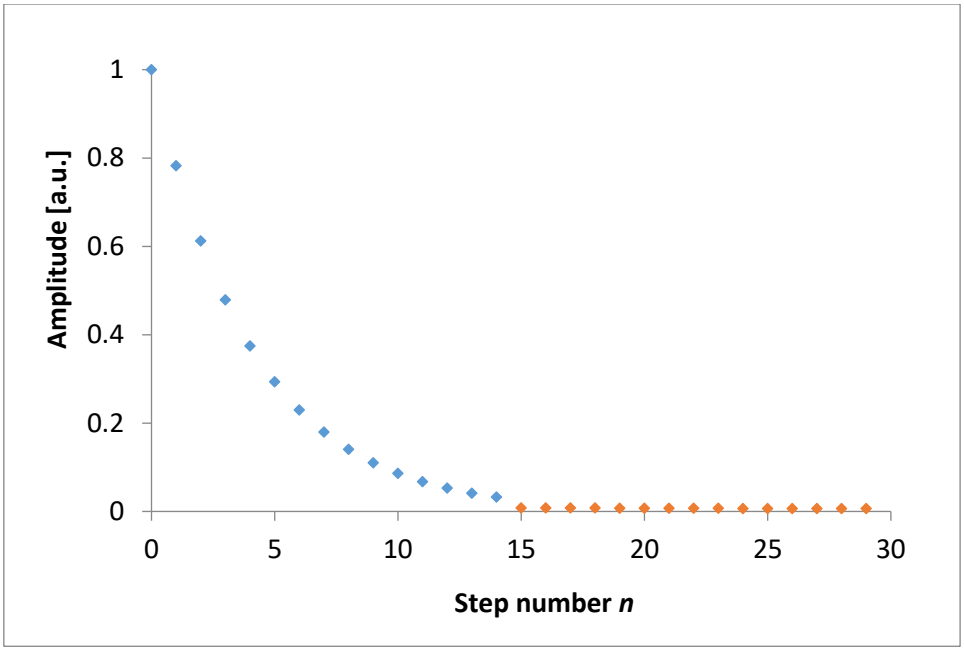


Figure 57. The signal amplitude at different nodes of the linear pathway for ω=0.4 s^{-1}, f=0.064 s^{-1}. The first 15 reactions are "slow" ($k_{1-15} = 0.5s^{-1}$), whereas the following 15 are "fast" ($k_{16-30} = 2s^{-1}$). A sharp discontinuity is observed where the rate constants change from slow to fast.

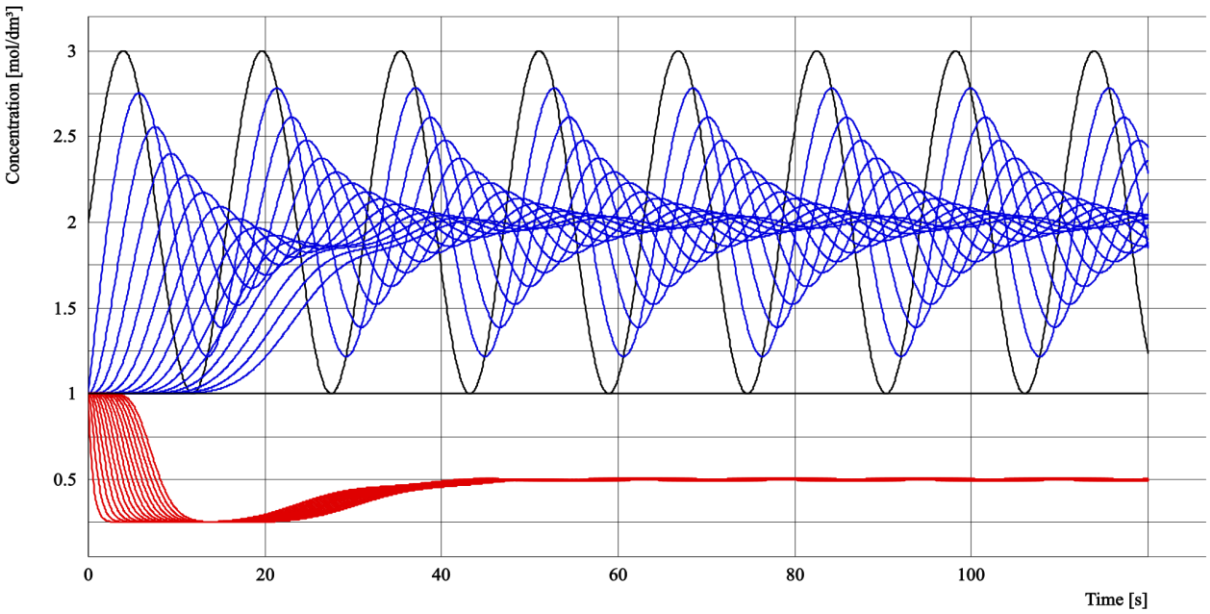


Figure 58. The concentration profile of the linear pathway for ω=0.4 s^{-1}, f=0.064 s^{-1}. The first 15 reactions are "slow" ($k_{1-15} = 0.5s^{-1}$), whereas the following 15 are "fast" ($k_{16-30} = 2s^{-1}$). The phase shifts and total concentrations for slow reactions are notably higher than those for fast reactions.

Next, I simulated the opposite situation – i.e., this time a "slow-to-fast" junction, where the consecutive 15 fast reactions were followed by 15 significantly slower ones:

$$a_0 \xrightarrow{fast} a_1 \xrightarrow{fast} a_2 \xrightarrow{fast} \dots \xrightarrow{fast} a_{13} \xrightarrow{fast} a_{14} \xrightarrow{fast} a_{15} \xrightarrow{slow} a_{16} \xrightarrow{slow} \dots \xrightarrow{slow} a_{28} \xrightarrow{slow} a_{29} \xrightarrow{slow} a_{30}$$

$$\frac{d[a_n]}{dt} = k_n[a_{n-1}], \qquad [a_0] = \sin 0.4t + 2, \qquad [a_{1-29}]_{init} = 1, \qquad [a_{30}] = const,$$

$$k_{1-15} = 2, \qquad k_{16-30} = 0.5\ [1/s].$$

This time, the signal at the junction was amplified – by as much as 312% – after which it gradually decayed along the chain, as shown in **Figures 59-60**. Apparently, oscillations can be strengthened from step to step if fast reactions are followed by sufficiently slow ones. This is an important discovery, showing that a signal could easily be amplified at any desired spot.

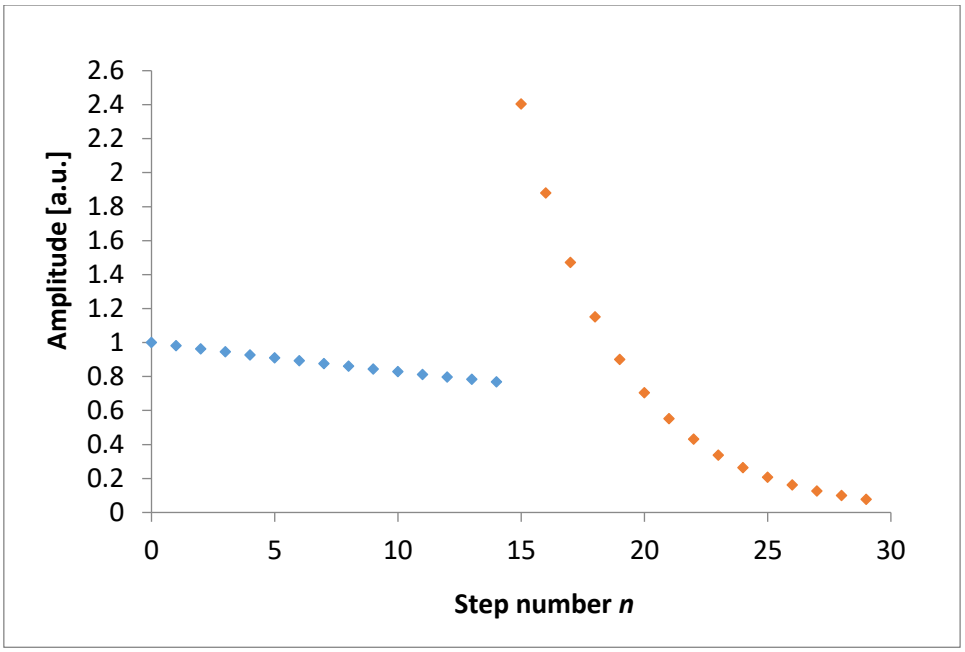


Figure 59. The signal amplitude at different nodes of the linear pathway for ω=0.4 s^{-1}, f=0.064 s^{-1}. The first 15 reactions are "fast" ($k_{1-15} = 2s^{-1}$), whereas the following 15 are "slow" ($k_{16-30} = 0.5s^{-1}$). A significant signal amplification is observed where the rate constants change from fast to slow.

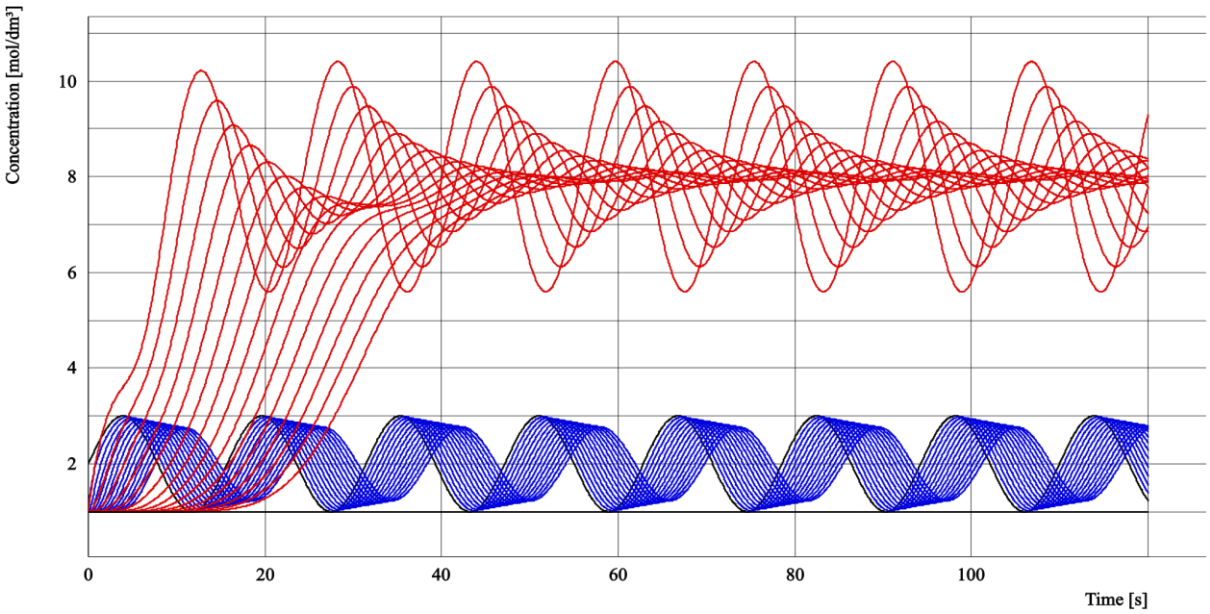


Figure 60. The concentration profile of the linear pathway for ω=0.4 s^{-1}, f=0.064 s^{-1}. The first 15 reactions are "fast" ($k_{1-15} = 2s^{-1}$), whereas the following 15 are "slow" ($k_{16-30} = 0.5s^{-1}$). The phase shifts and total concentrations for slow reactions are notably higher than those for fast reactions.

In order to assess this amplification effect and investigate the variables involved, I prepared the simplest possible model of a linear network, comprising only three reactions (inflow, transformation and outflow):

$$a_0 \xrightarrow{k_1=1} a_1 \xrightarrow{k_2=?} a_2$$

$$\frac{d[a_n]}{dt} = k_n[a_{n-1}], \qquad [a_0] = \sin \omega t + 2, \qquad [a_1]_{init} = 1, \qquad [a_2] = const,$$

$$\omega = 0.1, 0.4, \qquad k_1 = 1, \qquad k_2 = 0.5 - 3\ [1/s].$$

I have tested this model against two frequencies $\omega = 2\pi f$ – the results are shown in **Figure 61**, which confirms that the change in the maximum amplitude is not independent of the signal frequency.

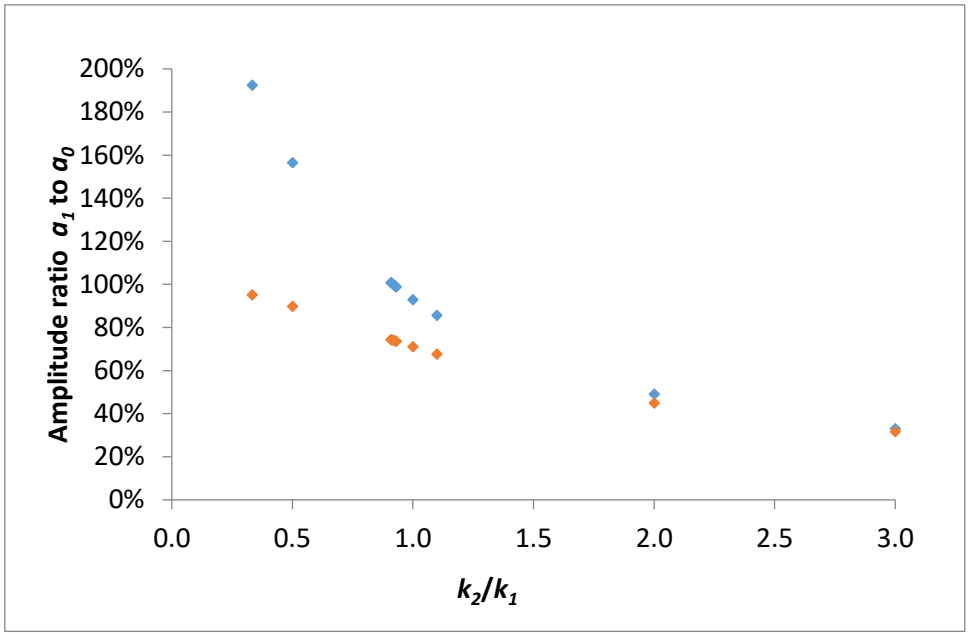


Figure 61. The amplitude decrease at varying ratios between the preceding and following reaction. The blue diamonds correspond to ω=0.4 s^{-1}, f=0.064 s^{-1}. The orange diamonds correspond to ω=0.1 s^{-1}, f= 0.016 s^{-1}.

As could be deduced from **Figure 61**, the relationship between the amplitude, frequency and reaction rates is not trivial. Fortunately, the equation could be found analytically and allows for an *a priori* prediction of the amplitudes.

Considering the specific case that resembles the previously described networks:

$$\frac{d[a_1]}{dt} = k_1[a_0] - k_2[a_1], \qquad [a_0] = [a_0]_A sin\omega t + [a_0]_{eq},$$

where $[a_0]_A$ is the maximum amplitude and $[a_0]_{eq}$ is the absolute concentration at times when the amplitude equals zero. Applying this description to a rate equation results in the following expression:

$$\frac{d[a_1]}{dt} = k_1\left([a_0]_A sin\omega t + [a_0]_{eq}\right) - k_2[a_1].$$

The solution of this differential equation is arguably long:

$$[a_1] = \frac{k_1}{{k_2}^2 + \omega^2}\left(k_2[a_0]_A sin\omega t - \omega[a_0]_A cos\omega t + \frac{[a_0]_{eq}\omega^2}{k_2} + [a_0]_{eq}k_2\right) + Ce^{-k_2 t},$$

however, in order to find the maximum amplitude of a_1, only the parts with trigonometric functions are needed. The remaining monomials can be omitted, giving the following expression of amplitude:

$$[a_1]_A = \frac{k_1[a_0]_A}{{k_2}^2 + \omega^2}(k_2 sin\omega t - \omega cos\omega t),$$

which can also be transformed into an expression of amplitude increase/decrease:

$$\frac{[a_1]_A}{[a_0]_A} = \frac{k_1}{{k_2}^2 + \omega^2}(k_2 sin\omega t - \omega cos\omega t).$$

In order to simplify this equation to only one trigonometric function, a certain assumption was made:

$$\frac{k_2}{\sqrt{{k_2}^2+\omega^2}} = cos\varphi, \qquad -\frac{\omega}{\sqrt{{k_2}^2+\omega^2}} = sin\varphi,$$

which with the following expression:

$$\frac{[a_1]_A}{[a_0]_A} = \frac{k_1}{\sqrt{{k_2}^2+\omega^2}}\left(\frac{k_2}{\sqrt{{k_2}^2+\omega^2}}sin\omega t - \frac{\omega}{\sqrt{{k_2}^2+\omega^2}}cos\omega t\right),$$

gives

$$\frac{[a_1]_A}{[a_0]_A} = \frac{k_1}{\sqrt{{k_2}^2+\omega^2}}(cos\varphi sin\omega t + sin\varphi cos\omega t).$$

According to the trigonometric identity of the sum of angles $sin(x+y) = sinxcosy + cosxsiny$:

$$\frac{[a_1]_A}{[a_0]_A} = \frac{k_1}{\sqrt{{k_2}^2+\omega^2}}sin(\omega t + \varphi),$$

but since only the maximum amplitude was measured in the simulations, the sine must be equal to 1. Thus, the general expression of amplitude decrease/increase at one reaction step is:

$$\frac{A_{n+1}}{A_n} = \frac{k_n}{\sqrt{{k_{n+1}}^2+(2\pi f)^2}}.$$

For the linear sequences comprising *m+1>2* reactions:

$$\frac{A_{n+1}}{A_n} = D_n, \qquad A_m = A_0 \cdot D_1 \cdot D_2 \cdot \ldots \cdot D_m, \qquad A_m = A_0 \prod_{n=1}^{m} D_n.$$

I have tested those equations against numerical solutions, with positive results. **Figure 62** compares the numerically calculated amplitudes, with the analytical solutions described – they match closely, with minimal deviations around the second or third decimal place. The exact values are given in **Table 1**.

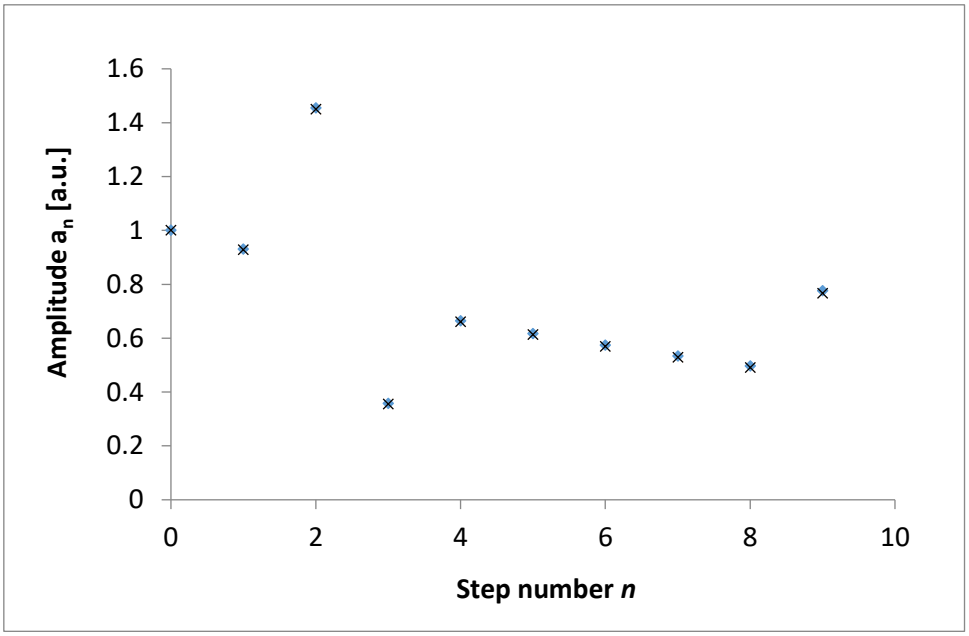


Figure 62. The signal amplitude at different nodes of the linear pathway, at ω=0.4 s^{-1}, f=0.064 s^{-1}, $k_1 = 1$, $k_2 = 1$, $k_3 = 0.5$, $k_4 = 2$, $k_5 = 1$, $k_6 = 1$, $k_7 = 1$, $k_8 = 1, k_9 = 1$, $k_{10} = 0.5\ [s^{-1}]$. The blue diamonds denote numerical solutions, whereas the black crosses denote analytical solutions. The exact values are given in **Table 1**.

Step *n*	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Numerical	0.999	0.930	1.455	0.357	0.664	0.617	0.574	0.533	0.496	0.776
Analytical	1	0.928	1.450	0.355	0.660	0.613	0.569	0.528	0.491	0.766

Table 1. The signal amplitude values presented in **Figure 62**.

The equation, where the amplitude increase/decrease after a sequence of reactions equals the product of the amplitude changes at each individual step, was also tested by creating a homogeneous chain – i.e., the rate constants at each step were all equal – and calculating how the amplitude on each consecutive step relates to its predecessor. As shown in **Figure 63** and **Table 2**, in such cases the amplitude decreases exponentially.

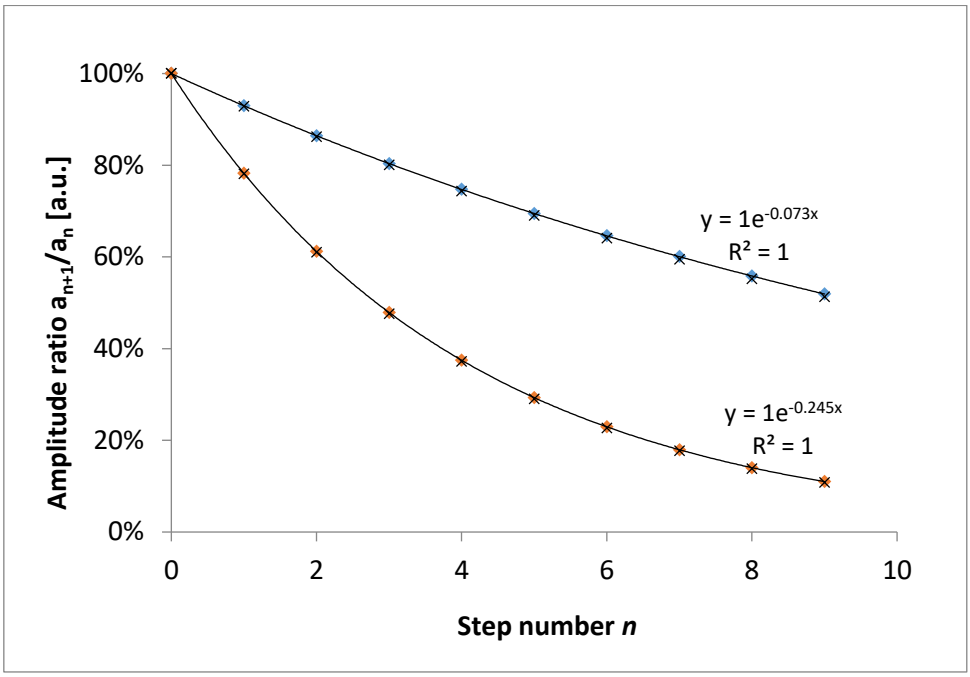


Figure 63. The amplitude decrease after a number *n* of subsequent steps, shown as a ratio between the amplitudes *n+1* and *n*, at frequency f=0.064 s^{-1}, ω=0.4 s^{-1}. The blue diamonds denote k_n=1 s^{-1}, whereas the orange diamonds denote k_n=0.5 s^{-1}.

k_n=1 s^{-1}

Step *n*	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Numerical	0.999	0.930	0.864	0.804	0.747	0.695	0.646	0.601	0.558	0.519
Ratio A_{n+1}/A_n	n/a	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930	0.930
Analytical	1	0.928	0.862	0.800	0.743	0.690	0.641	0.595	0.552	0.513

k_n=0,5 s^{-1}

Step *n*	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Numerical	0.999	0.782	0.612	0.479	0.375	0.293	0.229	0.179	0.140	0.110
Ratio A_{n+1}/A_n	n/a	0.782	0.782	0.782	0.782	0.782	0.782	0.782	0.782	0.782
Analytical	1	0.781	0.610	0.476	0.372	0.290	0.227	0.177	0.138	0.108

Table 2. The signal amplitude values from the plots in **Figure 63**. Note the constant **A_{n+1}/A_n** ratio.

The amplifying effect caused by the "fast-to-slow" transition suggests that the signal could be transferred along lengthy reaction sequences, while maintaining the initial strength, even at non-transmissive frequencies. However, the actual attempts to sustain such a signal by making each consecutive reaction slower allowed to carry the signal for up to only seven steps, after then it inevitably declined to almost zero. As shown in **Figure 64**, the rate decrease necessary to maintain the same maximum amplitude at every step was linear until the 4th reaction, after which preserving the initial amplitude required progressively lower rates. After the 7th step it was impossible to prevent the signal from dampening rapidly. In conclusion, the signals of non-transmissive frequencies cannot be propagated indefinitely through linear reaction networks.

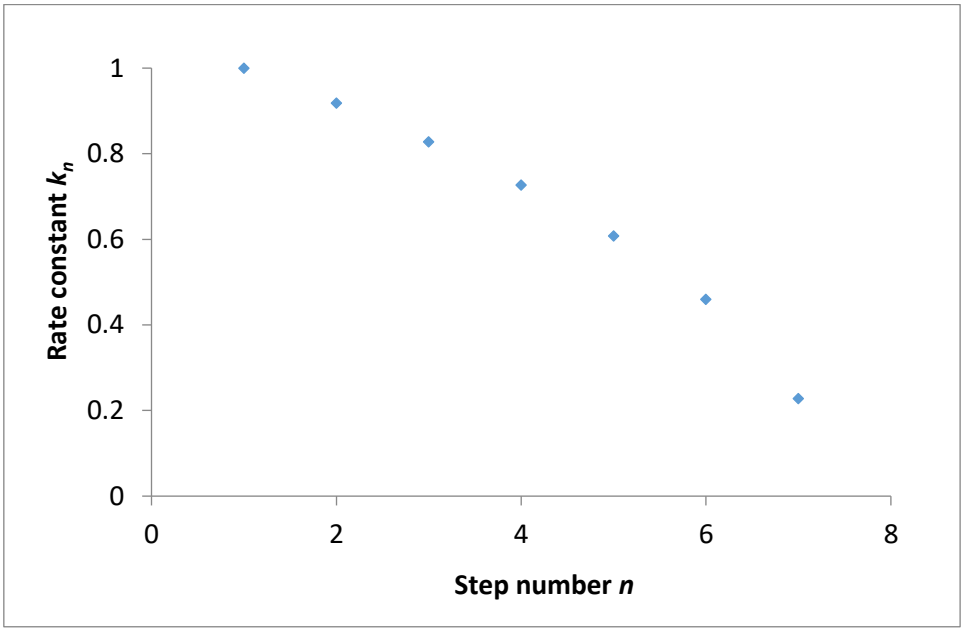


Figure 64. The rate constants of the subsequent reactions *n* required to maintain the unchanged signal strength, at ω=0.4 s^{-1}, f=0.064 s^{-1}. It was impossible to keep the amplitude unchanged after 7 steps, as it inevitably decreases.

In order to explain why the amplification effect does not allow to sustain the signal over long ranges, I performed a simulation where I could see how the signal transfer is affected by just one slow "interruption" in the middle of a chain. Predictably, the result is two discontinuities – strong amplification, then decrease – with a central "hot spot" in between. The lasting amplitude decrease in all the steps following afterwards was much less expected.

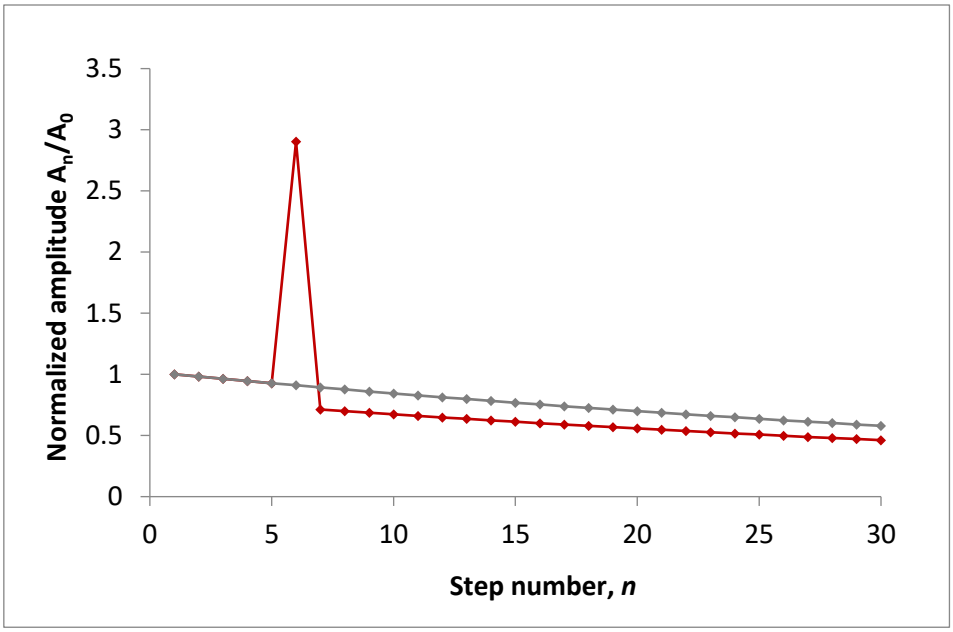


Figure 65. The signal amplitude at different nodes of the linear pathway for ω=0.4 s^{-1}, f=0.064 s^{-1}, denoted by the red plot. All the reactions are "fast" ($k_{1-5,\ 7-30} = 2s^{-1}$) except for one "slow" ($k_6 = 0.5s^{-1}$) at n=6. A significant amplitude "spike" is observed at the interrupting spot. The grey plot shows the same pathway without the "slow" reaction ($k_{1-30} = 2s^{-1}$) – a comparison of the two plots shows that the insertion of the "slow" step has effectively lowered the amplitudes of all the following steps.

Further simulations show that the position n of the interrupting "slow" step does not affect the long-term decrease (see **Figure 66**). The number of such steps, on the other hand, does (see **Figure 67**). Each consecutive "interruption" effectively shortens the "range" of the oscillatory signal in chemical networks – therefore it would be most effective to amplify the signal only once at its destination. This also explains why the attempts at sustained signal (from **Figure 64**) led to such a rapid decline after the 7th step.

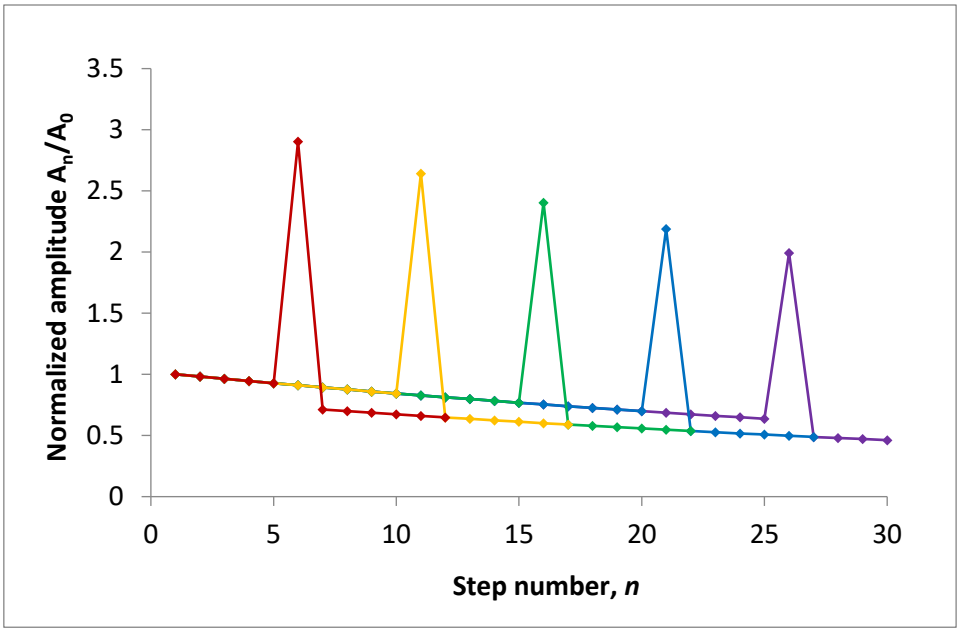


Figure 66. The signal amplitude at different nodes of the linear pathways for ω=0.4 s^{-1}, f=0.064 s^{-1}. All the networks comprise "fast" ($k_n = 2s^{-1}$) reactions plus one "slow" ($k_i = 0.5s^{-1}$) reaction at some specific spot *n*. The red plot denotes the sequence interrupted at *n*=6, the yellow at *n*=11, the green at *n*=16, the blue at *n*=21 and the purple at *n*=26. The position of the "slow" step is insignificant to long-term effects.

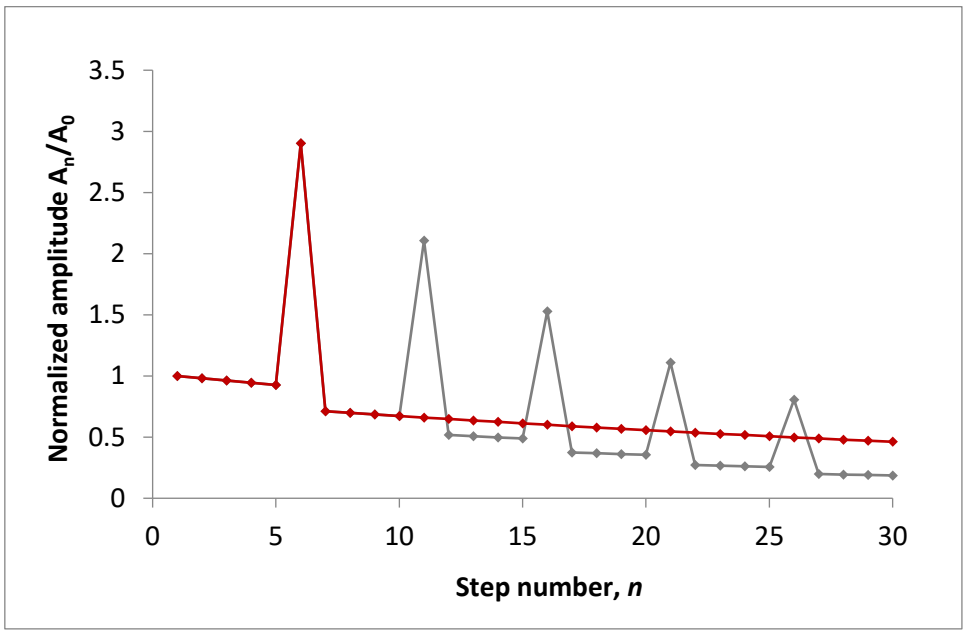


Figure 67. The signal amplitude at different nodes of the linear pathway for ω=0.4 s^{-1}, f=0.064 s^{-1}. The red plot comprises "fast" reactions ($k_n = 2s^{-1}$) plus one "slow" ($k_6 = 0.5s^{-1}$) at *n*=6. The grey plot shows a pathway with five "slow" reactions, located at *n*=6,11,16,21,26. Despite the immediately following amplifications, the more slow steps result in a stronger long-term dampening effect.

5.2.2 Predicting transition frequency

The initial experiment described in section **5.1** showed that every reaction (and, presumably, every chemical system) could be characterized as working similarly to a low-pass filter – i.e., eliminating signals above a certain "cut-off" frequency, while being fully transmissive to lower frequencies. The particular frequency varies between systems, often significantly. In just the simplest case, i.e., a single 1st order reaction followed by an outlet, which enables the pseudo-stationary state:

$$a_0 \rightarrow a_1 \rightarrow a_2,$$

$$\frac{d[a_n]}{dt} = k_n[a_{n-1}], \qquad [a_0] = \sin(2\pi f t) + 2, \qquad [a_1]_{init} = 1, \qquad [a_2] = const,$$

$$k_1 = 0.1 - 10, \qquad k_2 = 0.5 - 4, \qquad f = 0.1 - 10\,[1/s],$$

the threshold frequency is equal to a rate constant k_1, as shown in **Figure 68**. Any following reactions (i.e., k_2, outlet) only affect the "stretch" of a transition domain (this is shown in **Figure 69**).

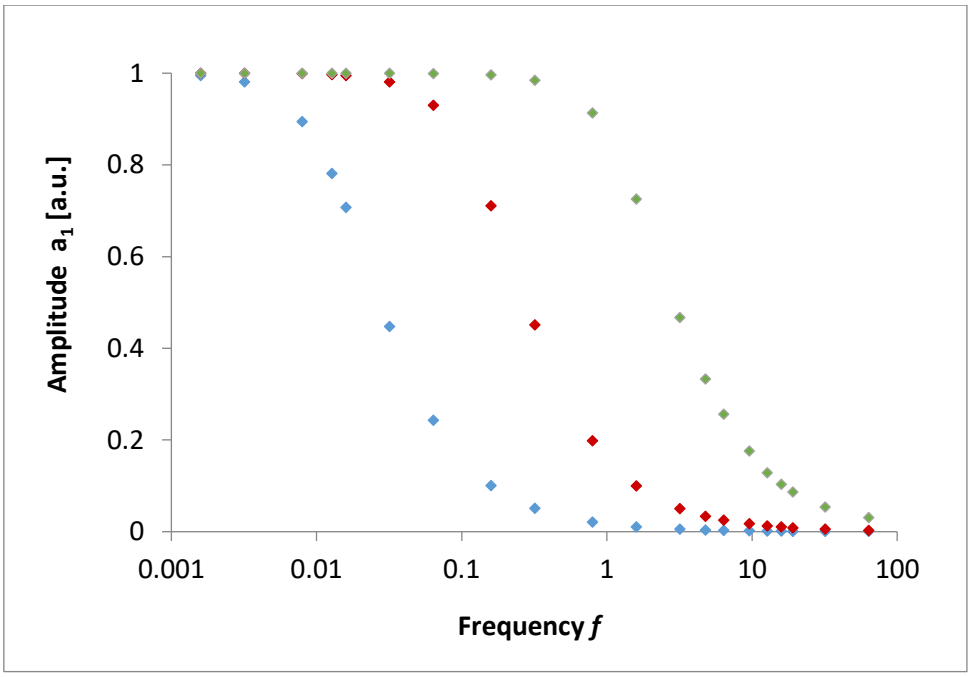


Figure 68. Strength of the propagated signal as a function of its frequency, measured as the concentration amplitude. The starting maximum amplitude a_0 is 1. The blue diamonds denote $k_1 = k_2 = 0.1s^{-1}$, the red – $k_1 = k_2 = 1s^{-1}$ and the green – $k_1 = k_2 = 10s^{-1}$. All the signals are transferred with ~15% of the initial strength at their respective threshold frequencies, equal to their rate constant k_1.

The threshold frequency is independent of the starting amplitude – the plots of A_1/A_0 versus $\frac{\omega}{k_1}$ fully overlap, even when A_0 varies. The plots in **Figure 69** show that the rate constant k_2, describing the output reaction, also does not affect the threshold frequency, but it does affect the resulting maximum amplitude A_1.

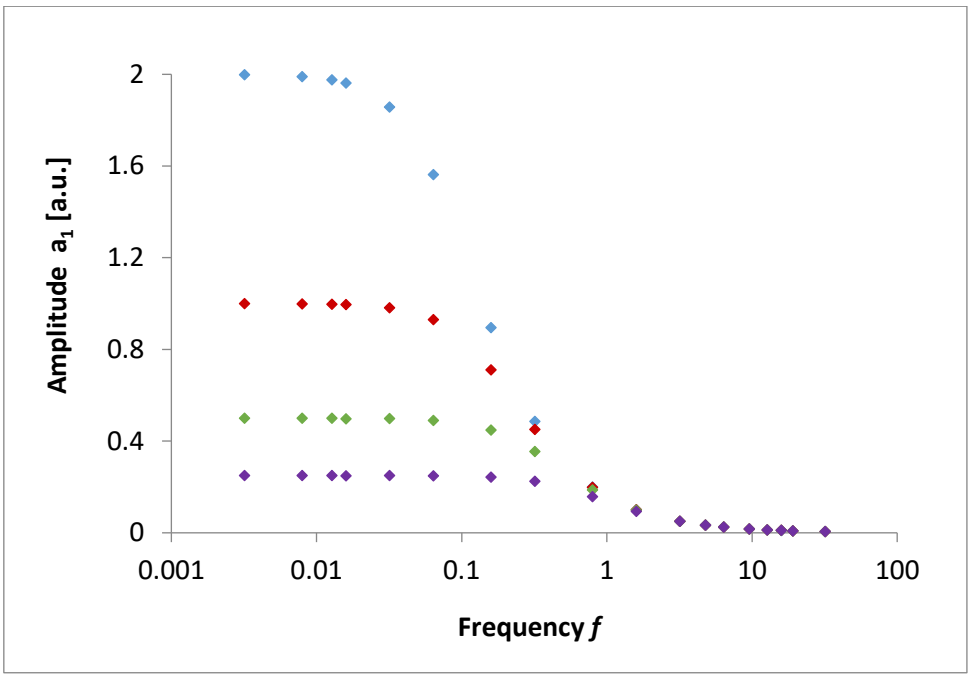


Figure 69. Strength of the propagated signal as a function of its frequency, measured as its concentration amplitude. The amplitude of a_0 is 1. In every case $k_1 = 1$. The blue diamonds denote $k_2 = 0.5s^{-1}$, the red – $k_2 = 1s^{-1}$, the green – $k_2 = 2s^{-1}$ and purple diamonds – $k_2 = 2s^{-1}$. The maximum strength of the transferred signal (observed at low frequencies) is $\sim\frac{k_1}{k_2}$. All the signals are transferred with ~15% of the maximum strength at their respective threshold frequencies, equal to their rate constant k_1.

This is an important observation – all the reactions strongly suppressed the oscillations at $\frac{f}{k} > 1$ – allowing for an *a priori* prediction whether a particular reaction will propagate the given oscillations or not – without the need for using Kinetix and numerical calculations. All it takes is to know the rate constants – apparent in single reactions, but not in entire chemical networks.

In case of the sequences comprising many reactions, the threshold frequency was not as apparent as the one of its parameters. The graph in **Figure 70** demonstrates the simplest case of a homogeneous linear network – i.e., one where all the reactions involved share the same rate constant *k*:

$$a_0 \rightarrow a_1 \rightarrow a_2 \rightarrow a_3 \rightarrow a_4 \rightarrow a_5 \rightarrow a_6 \rightarrow a_7 \rightarrow a_8 \rightarrow a_9 \rightarrow a_{10},$$

$$\frac{d[a_n]}{dt} = k_n[a_{n-1}], \qquad [a_0] = \sin \omega t + 2, \qquad [a_{1-9}]_{init} = 1, \qquad [a_{10}] = const,$$

$$k_1 = k_2 = k_3 = \cdots = k.$$

Albeit the plots of amplitude vs $\frac{f}{k}$ overlapped in **Figure 70**, they did not do so at $\frac{f}{k} \approx 1$. As visually appealing as it was, this overlap gave no indication of the threshold frequencies and was not observed when different steps *n* were compared (i.e., the chains of varying length did not have overlapping plots), as shown in **Figure 71**. Therefore, in the case of networks, we are unable to find the threshold frequencies by simply referring to the rate constants *k* even if all reactions share the same *k*.

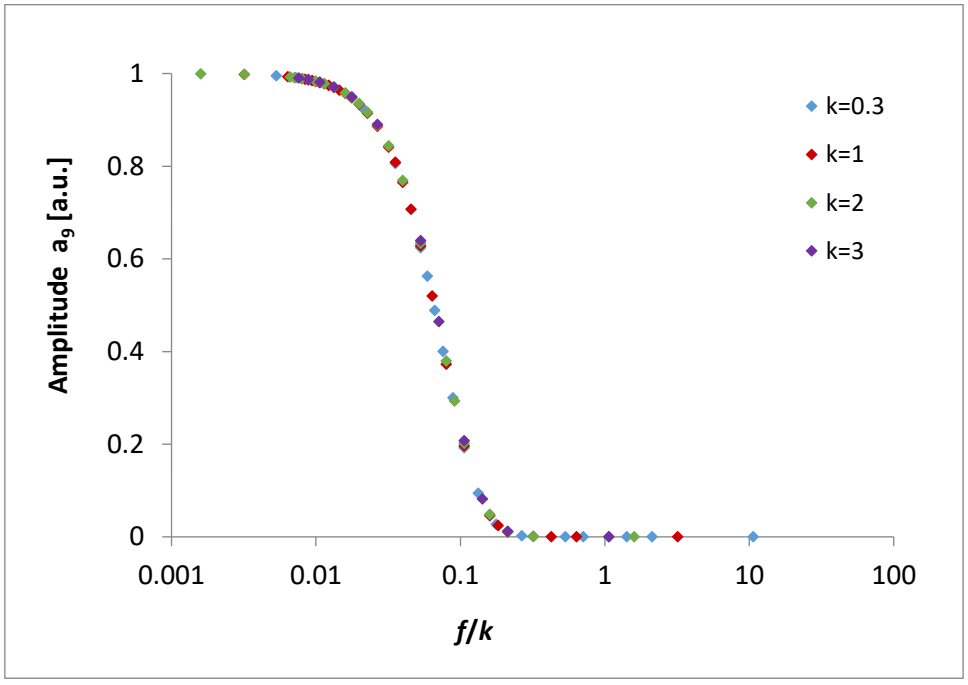


Figure 70. Strength of the propagated signal as a function of $\frac{f}{k}$, measured as a_9 concentration amplitude. The amplitude of a_0 is 1. All the steps share the same rate constant *k* – dividing frequency by this *k* results in overlapping of all the plots. The amplitude drops to ~15% at $\frac{f}{k} \approx 0.1$ instead of the predicted value of 1.

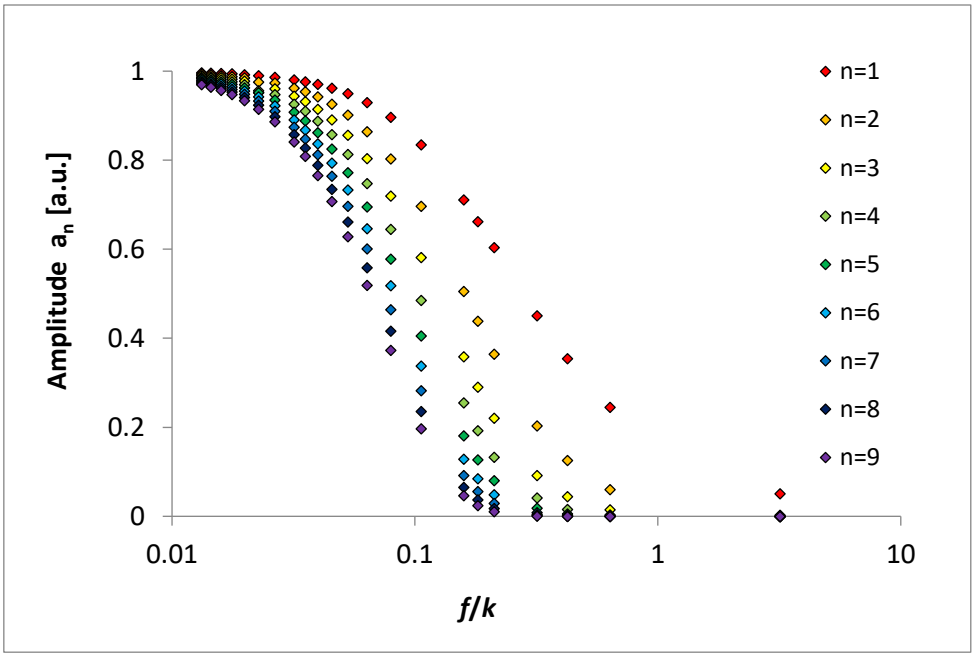


Figure 71. Strength of the propagated signal as a function of $\frac{f}{k}$, measured as the a_n concentration amplitude. All the steps share the same rate constant $k = 1$. The amplitude of a_0 is 1. For every *n*-numbered step, amplitude drops to ~15% of the starting value at different $\frac{f}{k}$.

If predicting the transmitted frequencies for single 1st order reactions *a priori* is possible, then predicting them for entire networks should also be possible. It would, however, require some "effective" rate constant comprising an entirety the system kinetics and architecture.

Referring to the expression for the reaction amplitude increase/decrease:

$$\frac{A_{n+1}}{A_n} = \frac{k_1}{\sqrt{{k_2}^2 + \omega^2}},$$

and to the equation regarding the chains of *m+1>2* reactions:

$$A_m = A_0 \prod_{n=1}^{m} \frac{A_{n+1}}{A_n},$$

allows for finding, analytically, the equation for the effective rate constant k_{eff} of entire 1st order reaction sequences.

Now, assuming that $\omega \ll k$ and $n \to \infty$, then for long times the expression:

$$\frac{A_n}{A_0} = \prod_{i=1}^{n} \frac{k_i}{\sqrt{{k_{i+1}}^2 + \omega^2}},$$

could be approximated to:

$$\frac{A_n}{A_0} = \left(\frac{k}{\sqrt{k^2 + \omega^2}}\right)^n = \frac{1}{\left(1 + \left(\frac{\omega}{k}\right)^2\right)^{\frac{n}{2}}} \approx \frac{1}{\sqrt{1 + 0.5n\left(\frac{\omega}{k}\right)^2}}.$$

Now, if the resulting expression is redefined as describing a single reaction instead of a full sequence:

$$0.5n\frac{1}{k^2} = \frac{1}{{k_{eff}}^2},$$

Then the expression becomes a formula for the effective rate constant of a linear pathway comprising *n* reactions:

$$\frac{1}{{k_{eff,n}}^2} = 0.5\sum_{i=1}^{n} \frac{1}{{k_{eff,i}}^2}.$$

This is an asymptotic approximation, for long and homogeneous chains. Nevertheless, when tested against the examples described earlier, the results were close to numerical solutions - all the oscillations for $\frac{f}{k_{eff}} > 1$ are completely dampened. **Figure 72** demonstrates this with the overlapping plots for varying *n*.

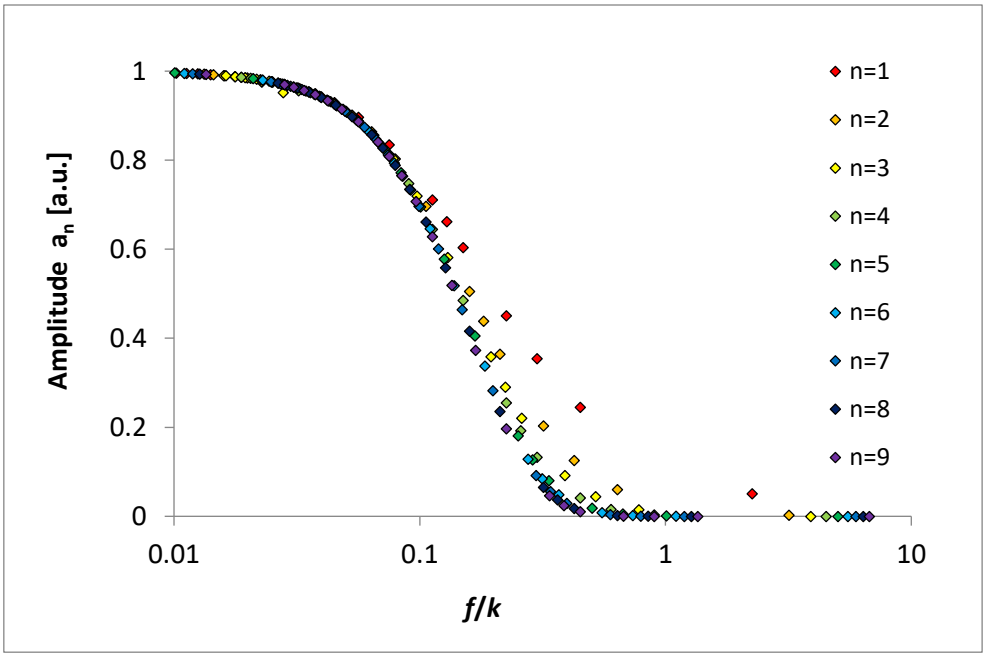


Figure 72. Strength of the propagated signal as a function of $\frac{f}{k_{eff}}$, measured as the a_n concentration amplitude. All the steps have the same rate constant $k = 1$. The amplitude of a_0 is 1. For $n \gg 1$ the plots overlap closely, while for $n = 1$ the plot deviates from the rest. Thus, the k_{eff} estimation method is suitable for longer sequences, while for a single reaction simply using its *k* gives a more accurate result.

The aforementioned equations also allow for an asymptotic analysis of extreme frequencies, far into the transmissive or the dampening domain. In the case of $\omega \ll k$ and for long times, the signal is transferred almost completely. As $\omega \to 0$, the amplitude decrease tends to:

$$\frac{A_n}{A_0} = \left(\frac{1}{1 + \left(\frac{\omega}{k}\right)^2} \right)^n,$$

which is well demonstrated in **Figure 73**.

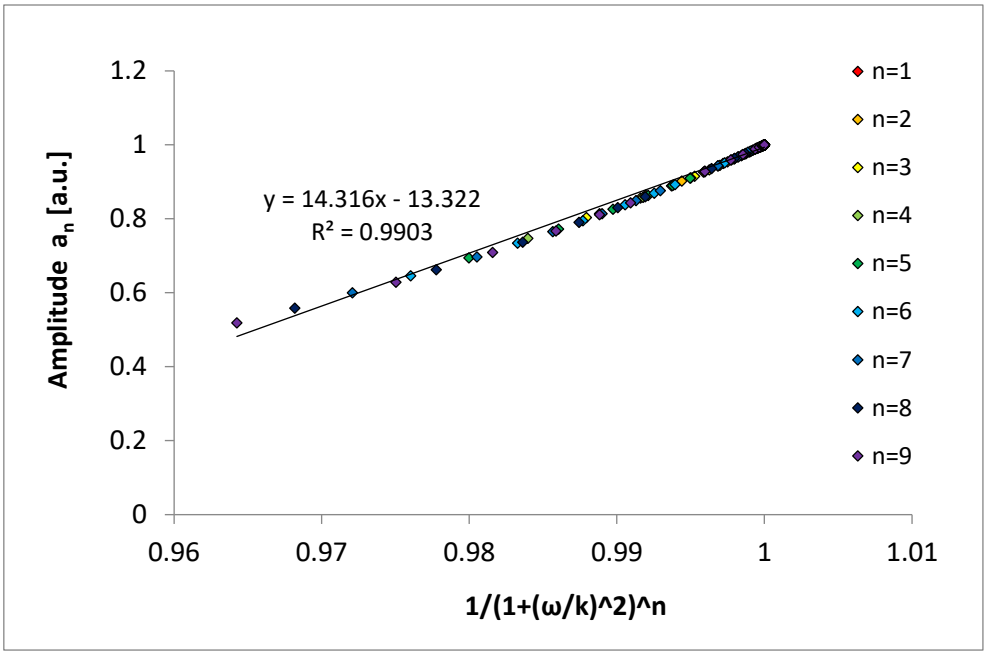


Figure 73. Strength of the propagated signal as a function of $\left(\frac{1}{1+\left(\frac{\omega}{k}\right)^2}\right)^n$, measured as the a_n concentration amplitude. All the steps have the same rate constant $k = 1$. The amplitude of a_0 is 1. The linear regression is made for n=9.

On the other hand, the signal was almost fully dampened at $\omega \gg k$. In such case, as $\omega \to \infty$, the amplitude decrease tends to:

$$\frac{A_n}{A_0} = \left(\frac{\frac{\omega}{k}}{1+\left(\frac{\omega}{k}\right)^2}\right)^n,$$

as shown in **Figure 74**.

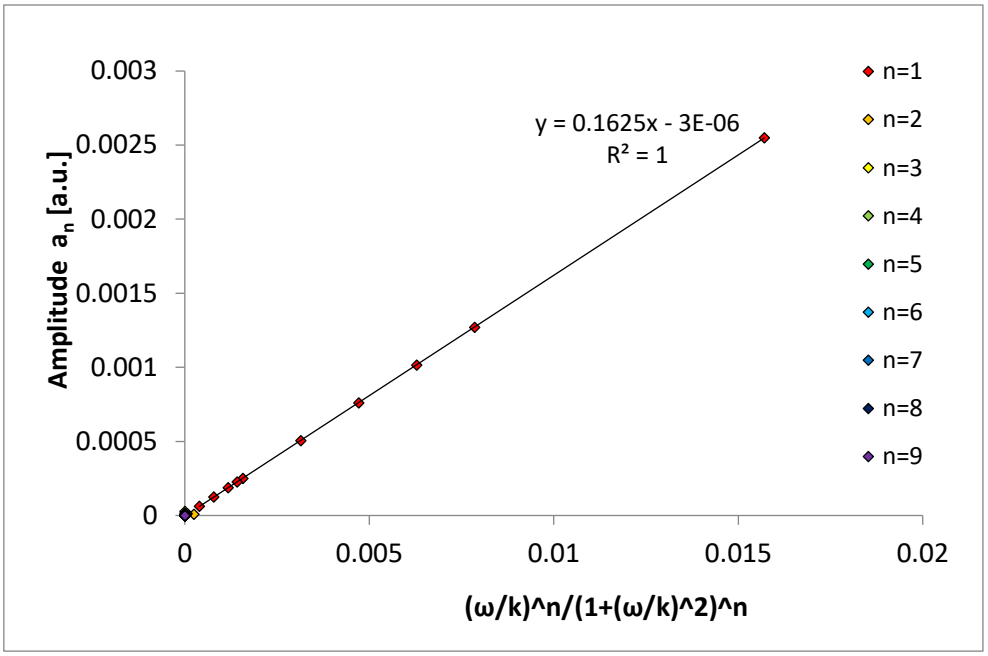


Figure 74. Strength of the propagated signal as a function of $\left(\frac{\frac{\omega}{k}}{1+\left(\frac{\omega}{k}\right)^2}\right)^n$, measured as the a_n concentration amplitude. All the steps have the same rate constant $k = 1$. The amplitude of a_0 is 1. The linear regression is made for n=1.

This ends the description of all three domains of signal propagation – the transmissive ($\omega \ll k$), the transition ($\omega \approx k$) and the dampening ($\omega \gg k$) domain – each can be localized in the frequency spectrum and their amplitude predicted *a priori*, without performing numerical calculations.

5.3. Propagation of signals through realistic reaction networks – part 1

Once the problem of the linear pathways had been solved, I began an investigation of more complex networks – specifically, kinase cascades. The model of a typical MAPK cascade presented in **Figure 75** features plausible reaction kinetics, but a simplified architecture. When initially designed and used by Kholodenko,[52] the model was an intentional simplification aimed at allowing for the analytical solution of diffusion equations. In reality, there are no known cascades with only one phosphorylation event in every layer (typically two are required for activation).

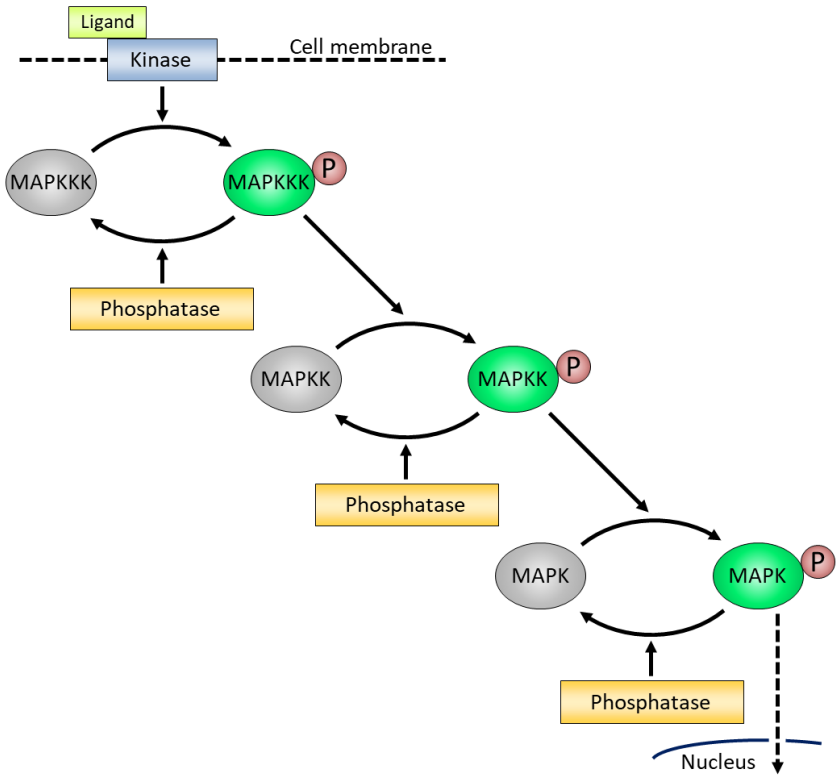


Figure 75. The model of a typical kinase cascade. The external signal acts on the membrane kinase (the blue square), which catalyzes the phosphorylation of MAPKKK (the grey ovals are inactive kinases, whereas the green ones are active/phosphorylated kinases). This phosphorylated MAPKKK-P then catalyzes the phosphorylation of MAPKK. The phosphatases (the orange squares) catalyze the dephosphorylation of the kinases – their concentrations will be constant, so the deactivation can be compared to a spontaneous process. The instantaneous concentration of MAPK-P can be considered a final output signal since it actively catalyzes the further processes occurring at the nucleus.

The catalytic activity of the kinases is described by the concentration of the kinase present in the rate equation. Mathematically speaking, in case of an example reaction:

$$a_0 \underset{c_2}{\overset{c_1}{\rightleftarrows}} a_1,$$

where *a* denotes reagents and *c* denotes catalysts, the rate equation would be:

$$\frac{d[a_1]}{dt} = [c_1] \cdot f(a_0), \qquad \frac{d[a_0]}{dt} = [c_2] \cdot f(a_1),$$

where – in the case of the following examples – $f(a_0)$ and $f(a_1)$ take form of the Michaelis-Menten rate equations. According to this scheme, the model can be represented with the following equations:

$$MAPKKK \underset{(2)}{\overset{stim\ (1)}{\rightleftarrows}} MAPKKK - P,$$

$$MAPKK \underset{(4)}{\overset{MAPKKK\ (3)}{\rightleftarrows}} MAPKK - P,$$

$$MAPK \underset{(6)}{\overset{MAPK\ (5)}{\rightleftarrows}} MAPK - P,$$

where the numbers in parentheses are used as a reference to the respective reactions in the following **Table 3**.

No.	Reaction	Rate expression	Kinetic constants
1	membrane kinase MAPKKK → MAPKKK-P	$v_1^{fw} = \frac{V_1^{fw} \cdot [MAPKKK]}{[MAPKKK] + K_{M1}}$	$V_1^{fw}/C = [stim]$ $K_{M1}/C = 0.5$
2	cytoplasmic phosphatase 1 MAPKKK-P → MAPKKK	$v_1^{rev} = \frac{V_1^{rev} \cdot [MAPKKK\_P]}{[MAPKKK\_P] + K_{Mp1}}$	$V_1^{rev}/C = 3\ \mathrm{s}^{-1}$ $K_{Mp1}/C = 0.7$
3	MAPKK kinase MAPKK → MAPKK-P	$v_2^{fw} = \frac{k_2^{fw} \cdot [MAPKKK\_P] \cdot [MAPKK]}{[MAPKK] + K_{M2}}$	$k_2^{fw} = 9\ \mathrm{s}^{-1}$ $K_{M2}/C = 0.7$
4	cytoplasmic phosphatase 2 MAPKK-P → MAPKK	$v_2^{rev} = \frac{V_2^{rev} \cdot [MAPKK\_P]}{[MAPKK\_P] + K_{Mp2}}$	$V_2^{rev}/C = 3\ \mathrm{s}^{-1}$ $K_{Mp2}/C = 0.7$
5	MAPK kinase MAPK → MAPK-P	$v_3^{fw} = \frac{k_3^{fw} \cdot [MAPKK\_P] \cdot [MAPK]}{[MAPK] + K_{M3}}$	$k_3^{fw} = 9\ \mathrm{s}^{-1}$ $K_{M3}/C = 0.5$
6	cytoplasmic phosphatase 3 MAPK-P → MAPK	$v_3^{rev} = \frac{V_3^{rev} \cdot [MAPKK\_P]}{[MAPKK\_P] + K_{Mp3}}$	$V_3^{rev}/C = 3\ \mathrm{s}^{-1}$ $K_{Mp3}/C = 0.7$

Table 3. The equations governing the cascade shown in **Figure 75**. The input signal is given by $[stim] = (\sin 2\pi ft + 1) \cdot 5\ \mathrm{nM}$ and total concentrations of the phosphorylated and unphosphorylated kinases at each layer are all equal, $C = 300\ \mathrm{nM}$. All the concentrations and Michaelis constants are expressed in nM.

The plots of amplitude vs frequency resemble those of the linear pathways, with a sharp transition between the transmissive and dampening domains – sharper for the final product, and wider for the intermediates, see **Figures 76-77.**

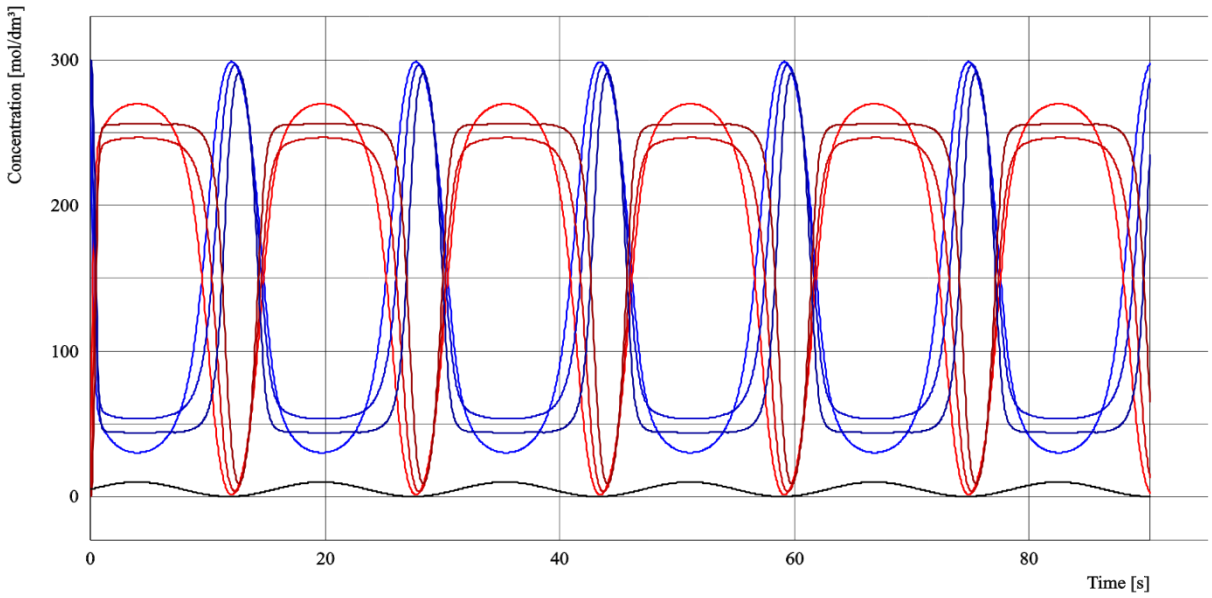


Figure 76. The concentration profiles of the kinase cascade described in **Figure 75** and **Table 3**. The black plot shows the input signal $[stim] = (\sin 2\pi f t + 1) \cdot 5$ nM. The blue plots describe inactive forms of the kinases, while the red ones describe active forms.

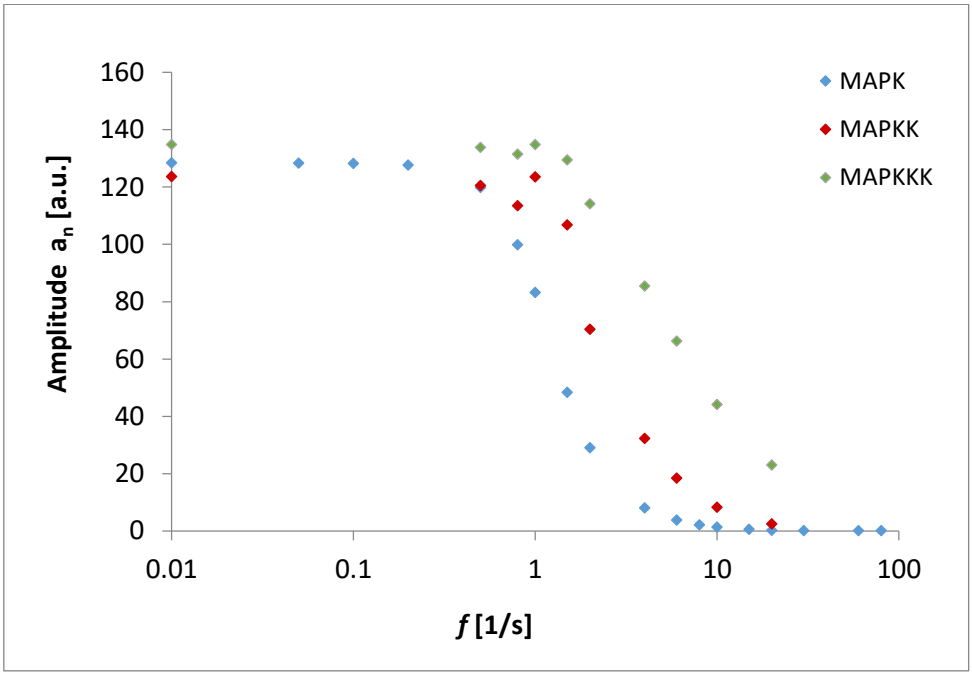


Figure 77. A response of the MAPK cascade to the oscillatory signals of various frequencies. The transition from the transmissive to dampening regimes for f greater than k_{eff} is similar to the linear pathways.

There is no apparent *k* in case of the enzymatic reactions with which to plot the curves against ω/k. The enzymatic rates are described with parameters, such as maximum velocity and Michaelis constants (intertwined in a non-linear manner), but not a classical rate constant like in linear equations.

Predictably, the analytical solutions regarding linear pathways do not apply to cascades – therefore, a new equation for k_{eff} had to be established, according to the analysis of the cascade network architecture.

5.4. Systematic analysis of non-linear reaction networks

This section describes step-by-step the methodology used for the analysis of non-linear systems. For the explicitly presented results, see section **6.**

The reaction networks referred to as "non-linear" in this work feature both a non-linear architecture and non-linear reaction rate equations. In order to reduce the number of variables during the analysis, both elements have been studied separately – first as non-linear kinetics in the linear networks (**5.4.1**), then as linear kinetics in the non-linear networks (**5.4.2**). Once both were fully understood, it was possible to analyze the model networks that are fully non-linear (as shown in sections **5.3.** and **5.5.**).

5.4.1. Higher-order and non-linear reaction kinetics

Before moving on to the cascading pathways, a basic problem had to be solved first – i.e., how to assess an effective rate constant from non-linear reaction kinetics. It was apparent in the case of the 1st order rate equations as it equaled the rate constant. In more complex cases, however, it had to be somehow estimated.

When searching for k_{eff}, I decided to treat the spontaneous deactivation of MAPK as analogous to the outflow reactions from the linear pathways. In case of the latter:

$$\frac{d[a_n]}{dt} = k_n[a_{n-1}] - k_{n+1}[a_n].$$

In this model, both k_n and k_{n+1} would be required to predict the amplitude changes. However, only k_n is necessary for predicting the transition frequencies if *n*=1, while k_{n+1} is not needed for any *n*. Analogically, for MAPK:

$$\frac{d[MAPK-P]}{dt} = v_3^{fw} - v_3^{rev} = \frac{k_3^{fw} \cdot [MAPKK_P] \cdot [MAPK]}{[MAPK] + K_{M3}} - \frac{V_3^{rev} \cdot [MAPKK_P]}{[MAPKK_P] + K_{Mp3}}.$$

This suggests that plotting $v_3^{fw}(t)$ – i.e., the gross increase in $[MAPK-P]$, instead of a net increase – would enable finding the reaction effective rate constant. In fact, for sufficiently long times a "forward rate" – $v_3^{fw}(t)$ (shown in **Figure 78**) – strongly resembles linear a 1st order reaction.

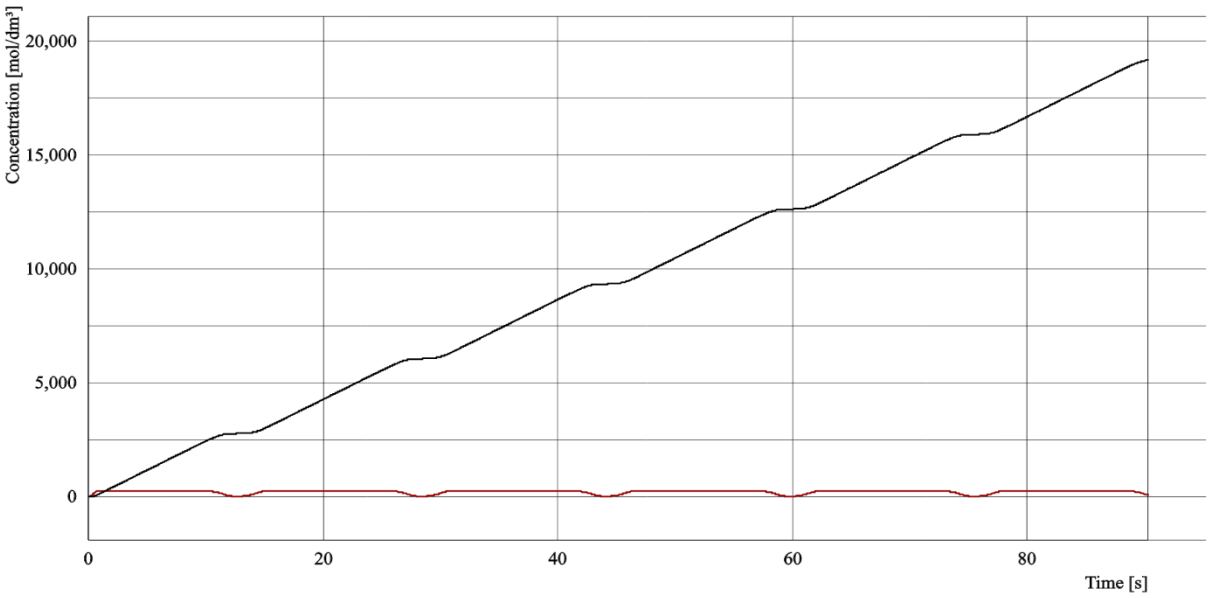


Figure 78. The concentration profile of MAPK-P. The red line denotes the net concentration of the active kinase. The black line denotes the gross concentration – i.e., theoretical rate of activation that disregards the simultaneous deactivation. As is presented, the black line can be approximated as the line for sufficiently long times and its slope as the estimation of the effective rate constant.

The linear regression of $v_3^{fw}(t)$ after a certain point in time t – i.e., the moment when all net concentration changes become periodic (also referred to as the "pseudo-stationary" state) – gives approximation of the "effective" rate constant *k* from its slope. In fact, in the realm of long times, where I measure all oscillations, every rate equation can be approximated as 1st order kinetics.

To show an example – the simplest Michaelis-Menten kinetics defined with equation:

$$\frac{d[N_i]}{dt} = \frac{V_{max}[N_{i-1}]}{[N_{i-1}] + K_M}, \qquad or \qquad \frac{d[N_i]}{dt} = [N_{i-1}] \cdot f([N_{i-1}]).$$

Since $f([N_{i-1}])$ is a non-linear function of $[N_{i-1}]$, it would typically be impossible to approximate to 1st order kinetics. However, if the substrate is fed systematically, then for $t \to \infty$ $f([N_{i-1}])$ can be approximated as a constant:

$$\frac{d[N_i]}{dt} = [N_{i-1}] \cdot f([N_{i-1}]) = [N_{i-1}] \cdot k_{eff,i},$$

allowing for estimation of the effective rate constant. Generally for all non-linear reactions transforming N_{i-1} into N_i:

$$\boldsymbol{k_{eff,i}} = \frac{d[N_i]}{dt} \cdot [N_{i-1}]^{-1}.$$

The same applies to reversible reactions, the simplest example being:

$$\frac{d[N_i]}{dt} = v_f - v_r = [N_{i-1}]k_f - [N_i]k_r.$$

Again, if the substrate is fed systematically, then for long reactions times:

$$[N_{i-1}]k_f - [N_i]k_r = [N_{i-1}] \cdot k_{eff,i},$$

and consequently:

$$\boldsymbol{k_{eff,i}} = k_f - k_r \frac{[N_i]}{[N_{i-1}]}.$$

Interestingly, the expression $\frac{[N_i]}{[N_{i-1}]}$ is independent of an input concentration, thus any state can be created to estimate $k_{eff,i}$ for sequences of this particular reaction type. This means that the ability of the system to transmit oscillations can be estimated from the numerical models of stationary state (i.e., in the absence of an input signal).

In order to verify the above-mentioned analysis, I have created reversible chains as test models (ended with one irreversible outflow reaction of the rate approximated to reversible steps):

$$a_0 \underset{\leftarrow}{\rightarrow} a_1 \underset{\leftarrow}{\rightarrow} a_2 \underset{\leftarrow}{\rightarrow} a_3 \underset{\leftarrow}{\rightarrow} a_4 \underset{\leftarrow}{\rightarrow} a_5 \underset{\leftarrow}{\rightarrow} a_6 \underset{\leftarrow}{\rightarrow} a_7 \underset{\leftarrow}{\rightarrow} a_8 \underset{\leftarrow}{\rightarrow} a_9 \rightarrow a_{10},$$

$$\frac{d[a_i]}{dt} = k_i^f[a_{i-1}] - k_i^r[a_i] - k_{i+1}^f[a_i], \qquad \frac{d[a_9]}{dt} = k_9^f[a_8] - k_9^r[a_9] - \left(k_{10}^f - k_{10}^r\right)[a_9],$$

$$\boldsymbol{k_{eff,i}} = k_i^f - k_i^r \frac{[a_i]}{[a_{i-1}]}, \qquad \frac{1}{{k_{eff,n}}^2} = 0.5 \sum_{i=1}^{n} \frac{1}{{k_{eff,i}}^2},$$

$$[a_0] = \sin \omega t + 2, \qquad [a_{1-9}]_{init} = 1, \qquad [a_{10}] = const,$$

$$k_1^f = k_2^f = k_3^f = \cdots = k_n^f = 1, \qquad k_1^r = k_2^r = k_3^r = \cdots = k_n^r = 0 - 0.75 \left[{}^1/_s\right],$$

and plotted them against frequency f (**Figure 79**) and $\frac{f}{k_{eff,n}}$ (**Figure 80**) for *n*=6.

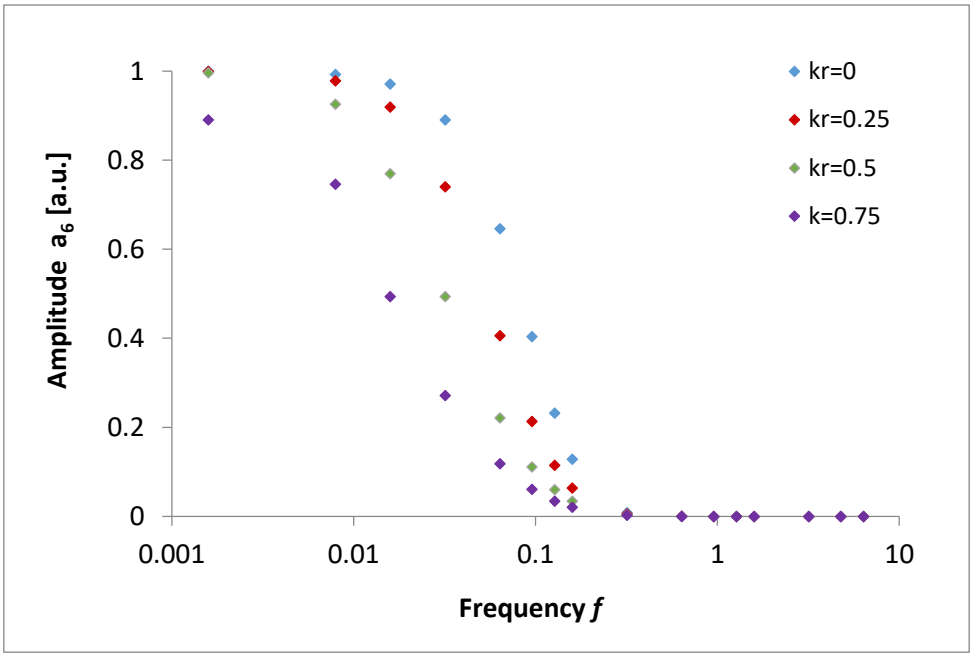


Figure 79. Strength of the signal propagated through a chain of reversible reactions as a function of its frequency, measured as the concentration a_6 amplitude. The amplitude of a_0 is 1. All the "forward" rate constants are equal to 1, whereas the "reverse" rate constants vary. The plots deviate slightly from each other.

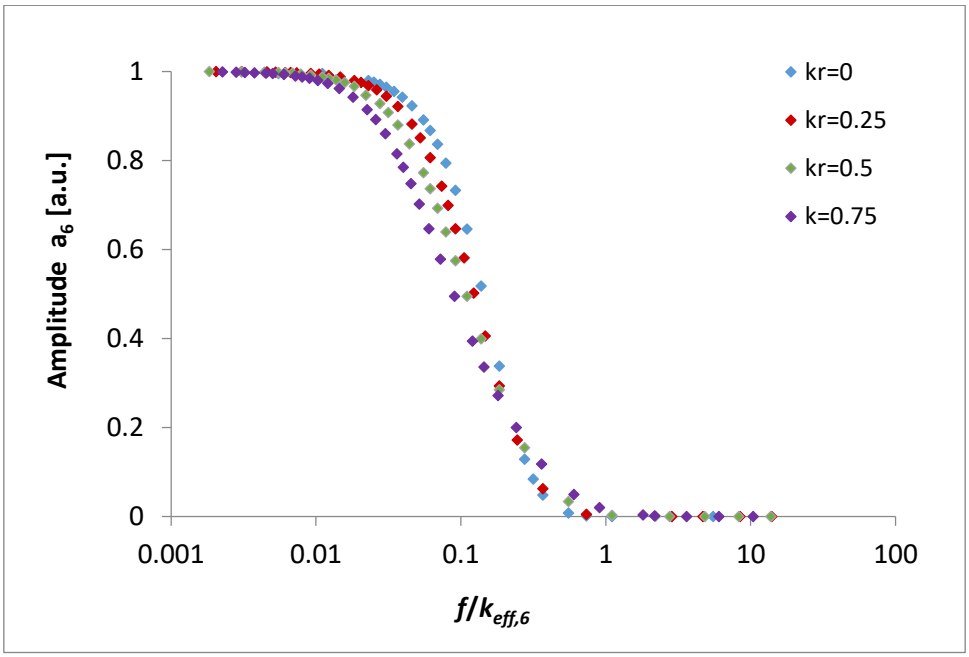


Figure 80. Strength of the signal propagated through a chain of reversible reactions as a function of $\frac{f}{k_{eff}}$, measured as the concentration a_6 amplitude. The amplitude of a_0 is 1. All the "forward" rate constants are equal to 1, whereas the "reverse" rate constants vary.

5.4.2 Predicting transition frequency in catalytic cascades

Although the equations of effective rate constant are available for linear pathways, they do not apply to MAPK since it has a completely different architecture. In order to study this type of network – a cascade – I have moved to a more accessible working model by reverting to 1st order kinetics and adding layers up to the total of 10. The resulting network is presented in **Figure 81**.

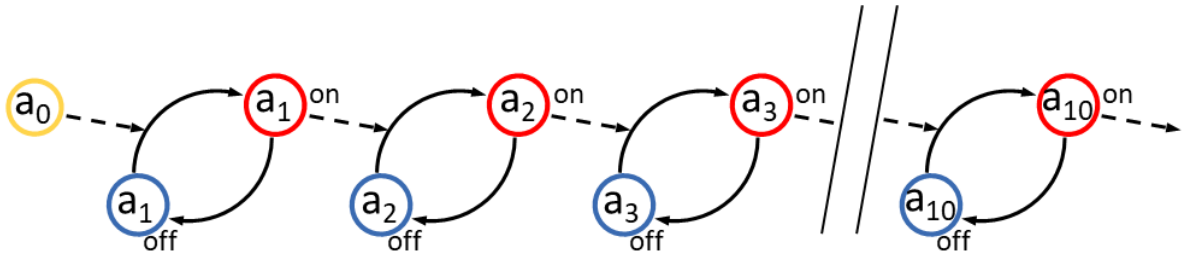


Figure 81. The scheme of a hypothetical catalytic cascade model. The input signal is denoted by the yellow circle, inactive forms of the catalyst – blue circles, active forms – red circles. There is a total of 10 layers – each comprising of an active and an inactive form assembled into an activation/deactivation loop. Activation at each layer is catalyzed by the active form of the previous layer, while deactivation is always spontaneous. The total concentration of both forms on each layer is always constant.

The model is, yet again, purely hypothetical, as it does not relate to any specific *in vivo* cascades:

$$a_1^{off} \underset{\leftarrow}{\overset{a_0}{\rightarrow}} a_1^{on}, \qquad a_2^{off} \underset{\leftarrow}{\overset{a_1^{on}}{\rightarrow}} a_2^{on}, \qquad a_3^{off} \underset{\leftarrow}{\overset{a_2^{on}}{\rightarrow}} a_3^{on}, \qquad a_4^{off} \underset{\leftarrow}{\overset{a_3^{on}}{\rightarrow}} a_4^{on}, \qquad a_5^{off} \underset{\leftarrow}{\overset{a_4^{on}}{\rightarrow}} a_5^{on},$$

$$a_6^{off} \underset{\leftarrow}{\overset{a_5^{on}}{\rightarrow}} a_6^{on}, \qquad a_7^{off} \underset{\leftarrow}{\overset{a_6^{on}}{\rightarrow}} a_7^{on}, \qquad a_8^{off} \underset{\leftarrow}{\overset{a_7^{on}}{\rightarrow}} a_8^{on}, \qquad a_9^{off} \underset{\leftarrow}{\overset{a_8^{on}}{\rightarrow}} a_9^{on}, \qquad a_{10}^{off} \underset{\leftarrow}{\overset{a_9^{on}}{\rightarrow}} a_{10}^{on},$$

$$a_n^{off} \underset{\leftarrow}{\overset{a_{n-1}^{on}}{\rightarrow}} a_n^{on}, \qquad \frac{d[a_n^{on}]}{dt} = k_n^{fw}[a_{n-1}^{on}][a_n^{off}], \qquad \frac{d[a_n^{off}]}{dt} = k_n^{rev}[a_n^{on}],$$

$$C_n = [a_n^{on}] + [a_n^{off}] = const, \qquad [a_0] = A_0 sin(\omega t) + B_0.$$

Unfortunately, this particular cascade had no analytical solutions and all the attempted approximations were prohibitively complex. The sought expressions had to be found with numerical methods, which were not particularly straightforward.

In case of the cascades, there were three variables for every layer – the total concentration C_n, the "forward" (k_{fw}) and "reverse" (k_{rev}) rate constants k – and the baseline of a_0 oscillations, B_0, i.e., the absolute concentration of a_0 at times when the amplitude equaled zero. In contrast to the linear pathways, B_0 does affect the outcome signal – this occurs because cascades do have an equilibrium point, whereas the linear pathways could only reach the stationary state. The problem is further simplified in the cascades where all layers are identical:

$$k_1^{fw} = k_2^{fw} = \cdots = k_n^{fw}, \qquad k_1^{rev} = k_2^{rev} = \cdots = k_n^{rev}, \qquad C_1 = C_2 = \cdots = C_n.$$

First, the relation between the signal propagation and the total concentration was investigated by performing a series of simulations with varying total concentration C. All the other parameters were constant and the input signal oscillated between 0 and 1:

$$k_1^{fw} = 1\ [a.u./s], \qquad k_1^{rev} = 0.5\ [1/s], \qquad A_0 = B_0 = 0.5\ [a.u.].$$

The results of these simulations are presented in **Figures 82-89**. In the time-concentration profiles, the most apparent was the unusual shape of the plots – as if the starting oscillations of $\sin x$-type were transformed into $e^{\cos x}$ upon transmission. Moreover, at low C, the average concentration of the inactive form was higher than the active one, but as C increases, the opposite became true.

In the plots of amplitude vs frequency, however, the most remarkable was the fact that the signal was never transmitted in full strength, even in the supposed transmissive domains – instead, presumably, the amplitude showed asymptotic behavior for $n \to \infty$. Lastly, the increase in C resulted in the larger "stretch" of the threshold domain.

The subsequent simulations, shown in **Figures 90-97**, allowed to establish that for $n \to \infty$, the signal baselines B_n asymptotically tended to $B_n = C - \frac{k_{rev}}{k_{fr}}$ – this was confirmed by a series of simulations, where $B_0 = C - \frac{k_{rev}}{k_{fr}}$ (shown in **Figures 98-103**). In those cases, the time-concentration profiles overlapped neatly, even at initial steps n. On the other hand, the simulations, where $B_0 \gg C - \frac{k_{rev}}{k_{fr}}$ and $A_0 \ll B_0$ (shown in **Figures 104-113**), displayed an apparent gradual shift in B_n towards $C - \frac{k_{rev}}{k_{fr}}$. This shift was more significant the greater the difference between B_0 and $C - \frac{k_{rev}}{k_{fr}}$.

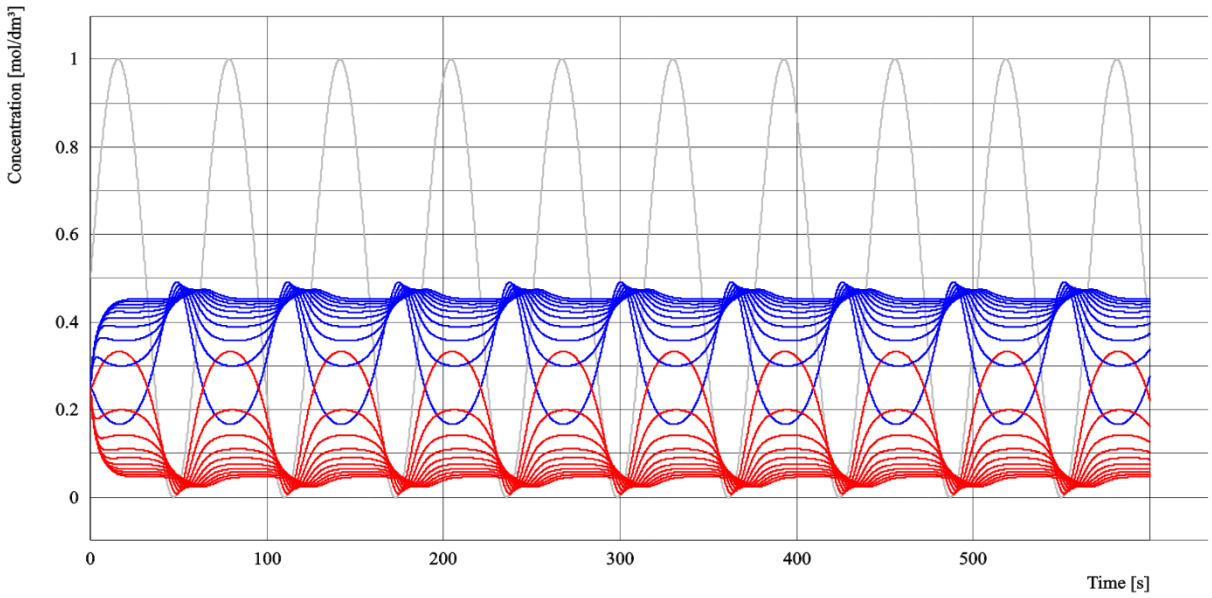


Figure 82. The concentration profile of a catalytic cascade. k_f **=1** u/s^{-1}, C **=0.5** u, k_r **=0.5** s^{-1}, A_0**=0.5** u, B_0 **=0.5** u, where u is a unit of concentration. The inactive form apparently tends to the concentration of $C - \frac{k_r}{k_f} = 0\ u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

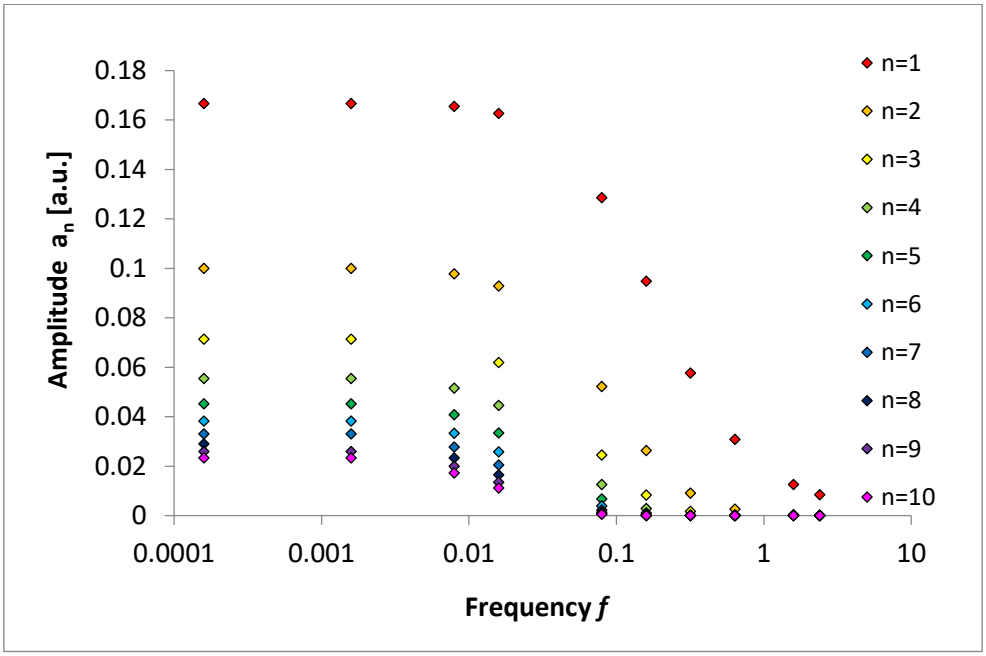


Figure 83. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the n-numbered active form. k_f **=1** u/s^{-1}, C **=0.5** u, k_r **=0.5** s^{-1}, A_0**=0.5** u, B_0 **=0.5** u, where u is a unit of concentration.

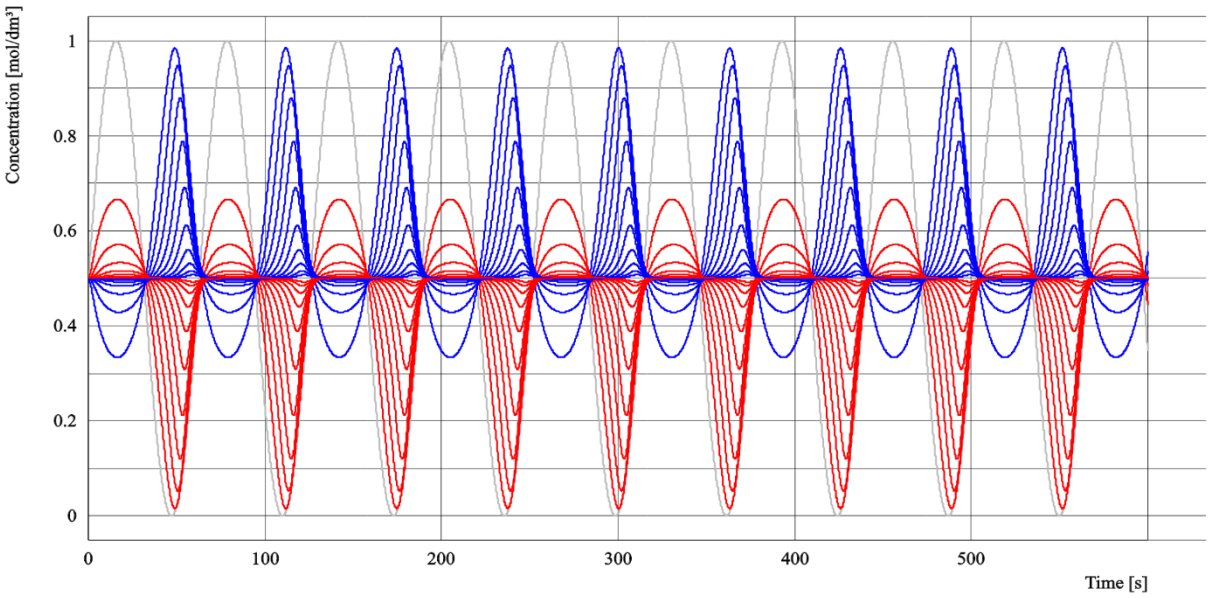


Figure 84. The concentration profile of a catalytic cascade. k_f **=1** ***u*****/s^{-1}**, ***C*** **=1** ***u***, k_r **=0.5 s^{-1}**, A_0**=0.5** ***u*****,** B_0 **=0.5** ***u***, where *u* is a unit of concentration. The inactive form apparently tends to the concentration of $C - \frac{k_r}{k_f} = 0.5\ u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

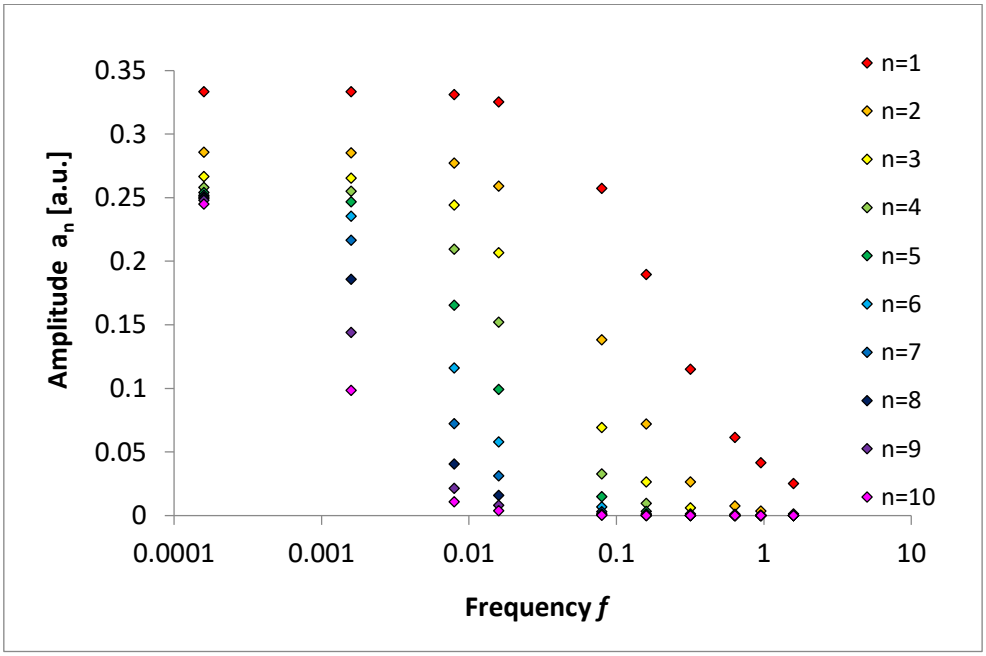


Figure 85. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f **=1** ***u*****/s^{-1}**, ***C*** **=1** ***u***, k_r **=0.5 s^{-1}**, A_0**=0.5** ***u*****,** B_0 **=0.5** ***u***, where *u* is a unit of concentration.

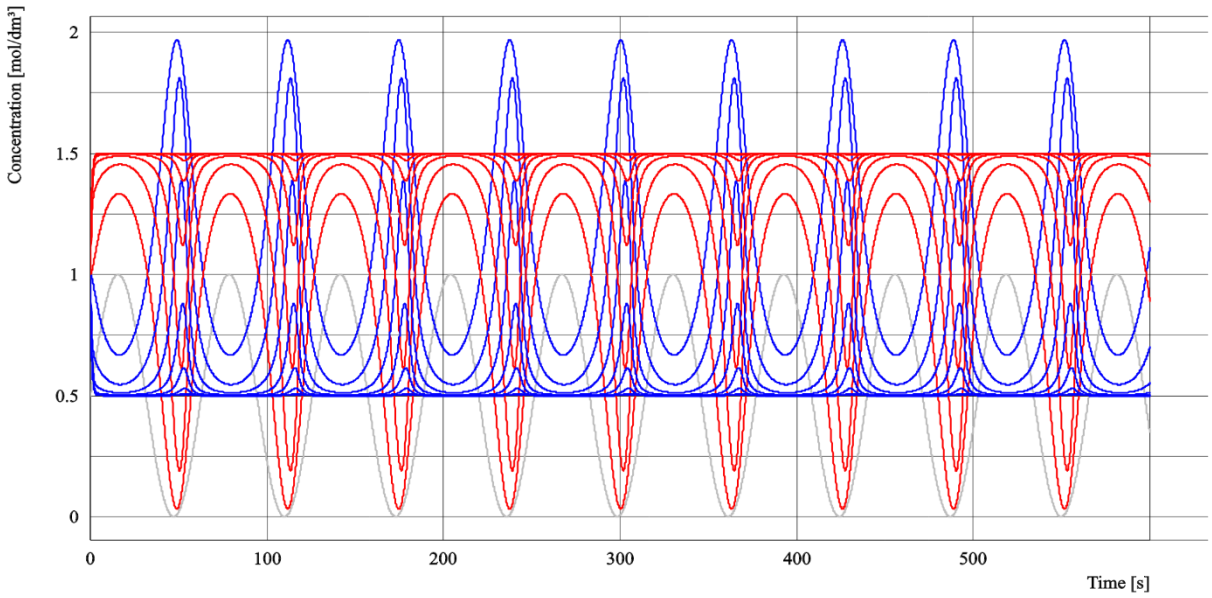


Figure 86. The concentration profile of a catalytic cascade. k_f **=1** u/s^{-1}, C **=2** u, k_r **=0.5** s^{-1}, A_0**=0.5** u, B_0 **=0.5** u, where u is a unit of concentration. The inactive form apparently tends to the concentration of $C - \frac{k_r}{k_f} = 1.5\ u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

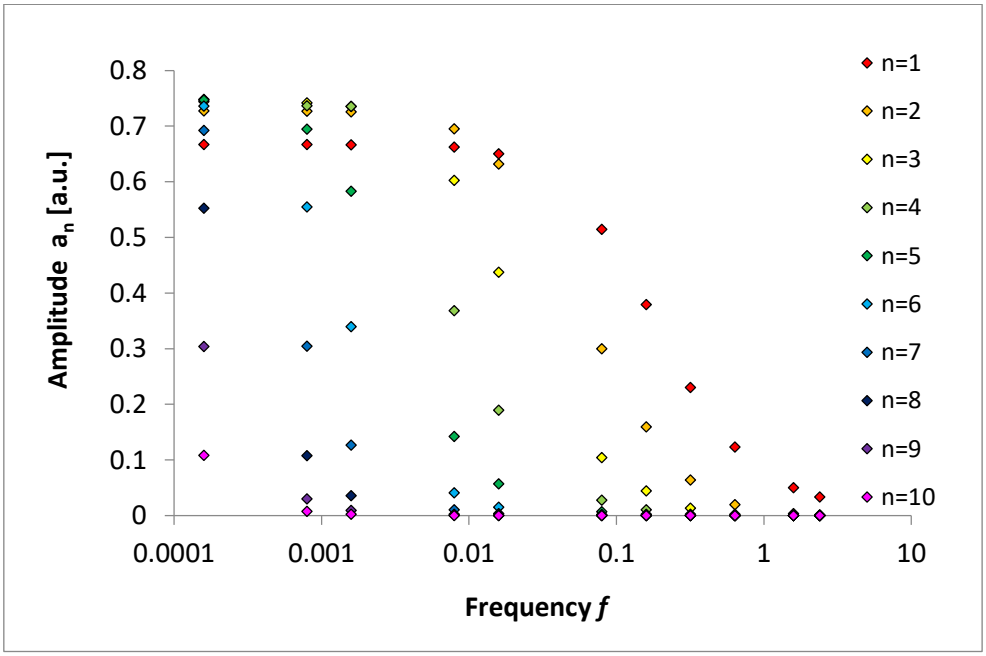


Figure 87. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the n-numbered active form. k_f **=1** u/s^{-1}, C **=2** u, k_r **=0.5** s^{-1}, A_0**=0.5** u, B_0 **=0.5** u, where u is a unit of concentration.

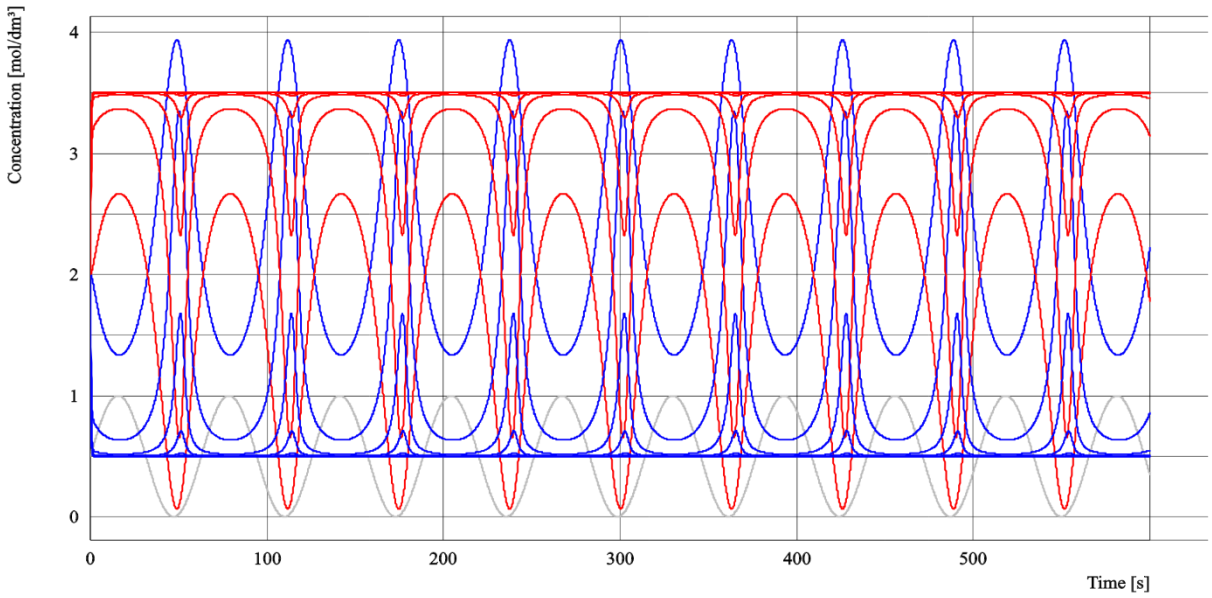


Figure 88. The concentration profile of a catalytic cascade. k_f **=1** u/s^{-1}, C **=4** u, k_r **=0.5** s^{-1}, A_0**=0.5** u, B_0 **=0.5** u, where u is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 3.5\ u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

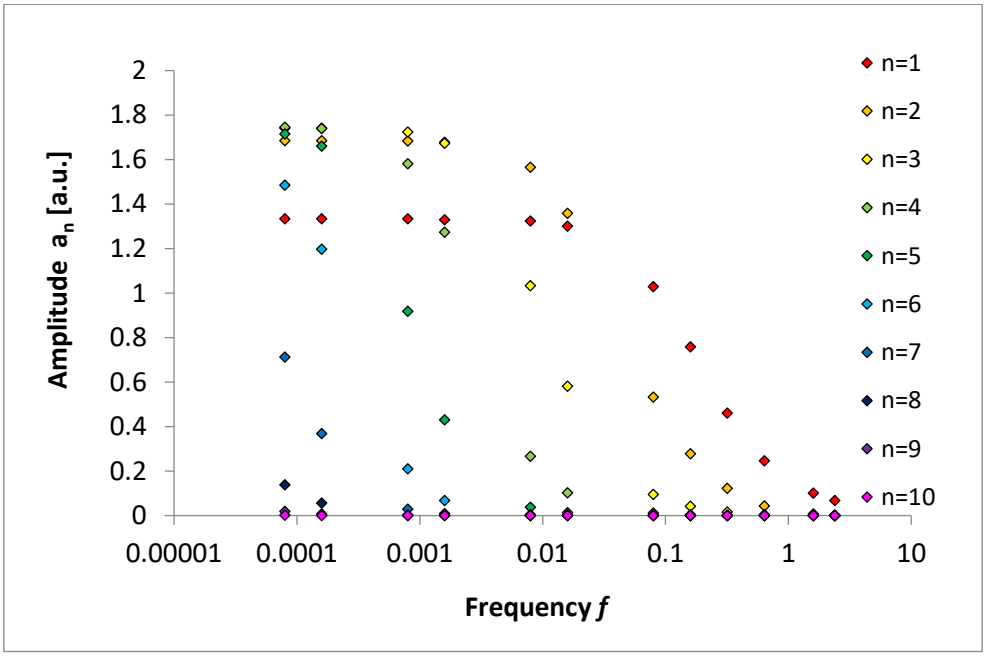


Figure 89. Strength of the signal propagated through a cascade a function of its frequency, measured as the amplitude of the n-numbered active form. k_f **=1** u/s^{-1}, C **=4** u, k_r **=0.5** s^{-1}, A_0 **=0.5** u, B_0 **=0.5** u, where u is a unit of concentration.

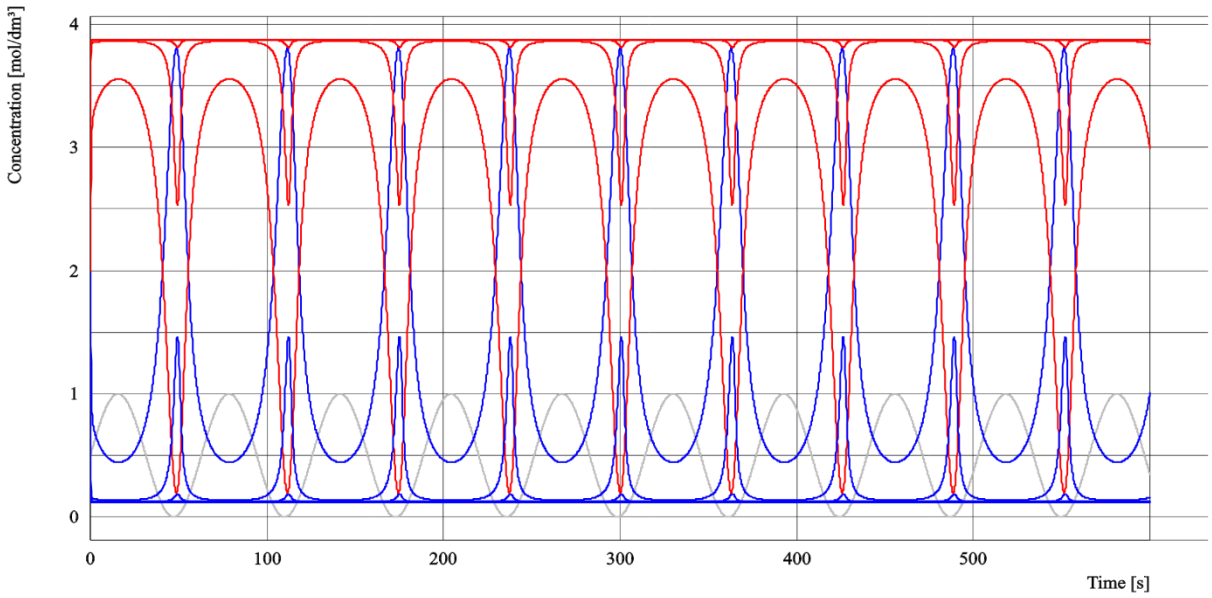


Figure 90. The concentration profile of a catalytic cascade. k_f =4 *u*/s^{-1}, *C* =4 *u*, k_r =0.5 s^{-1}, A_0=0.5 *u*, B_0 =0.5 *u*, where *u* is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 3.875\ u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

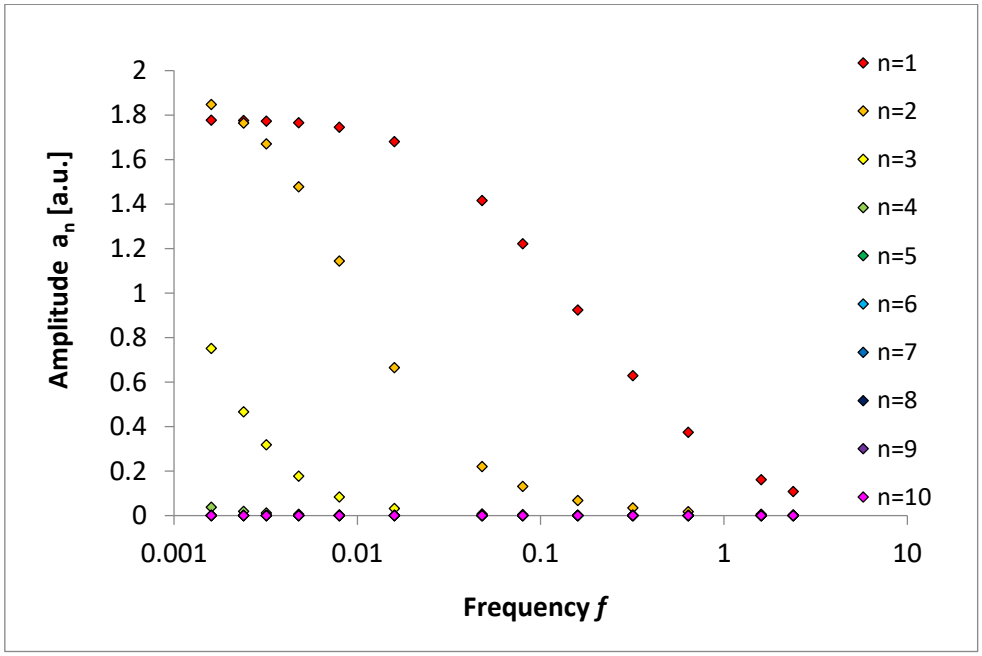


Figure 91. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f =4 *u*/s^{-1}, *C* =4 *u*, k_r =0.5 s^{-1}, A_0=0.5 *u*, B_0 =0.5 *u*, where *u* is a unit of concentration.

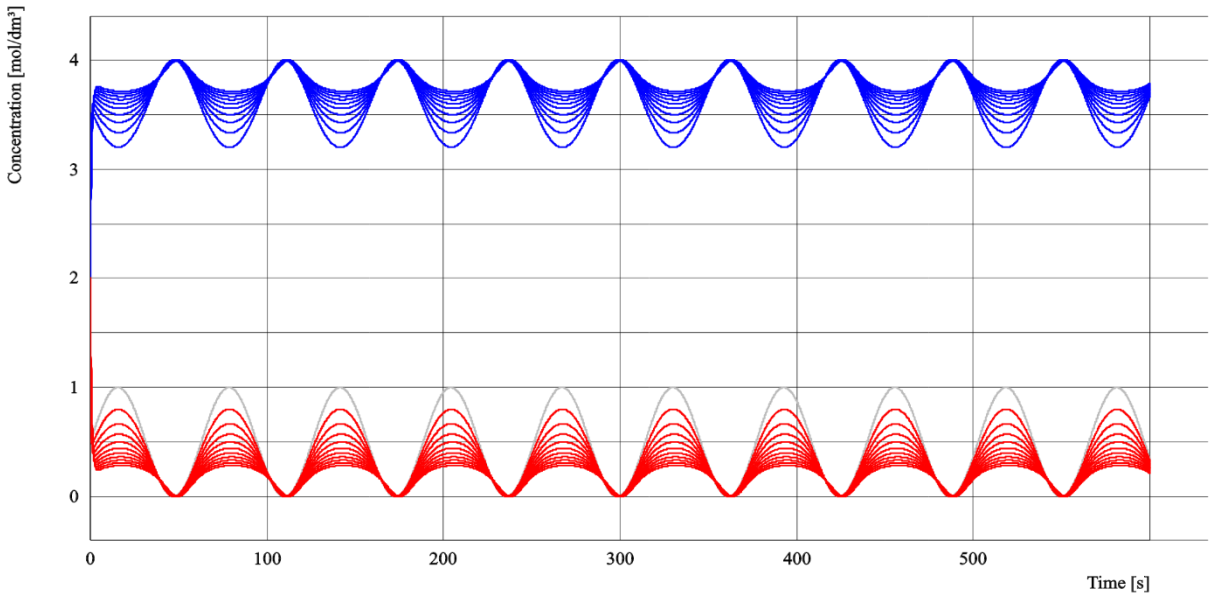


Figure 92. The concentration profile of a catalytic cascade. k_f=1 *u*/s^{-1}, C=4 *u*, k_r=4 s^{-1}, A_0=0.5 *u*, B_0=0.5 *u*, where *u* is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 0$ *u* with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

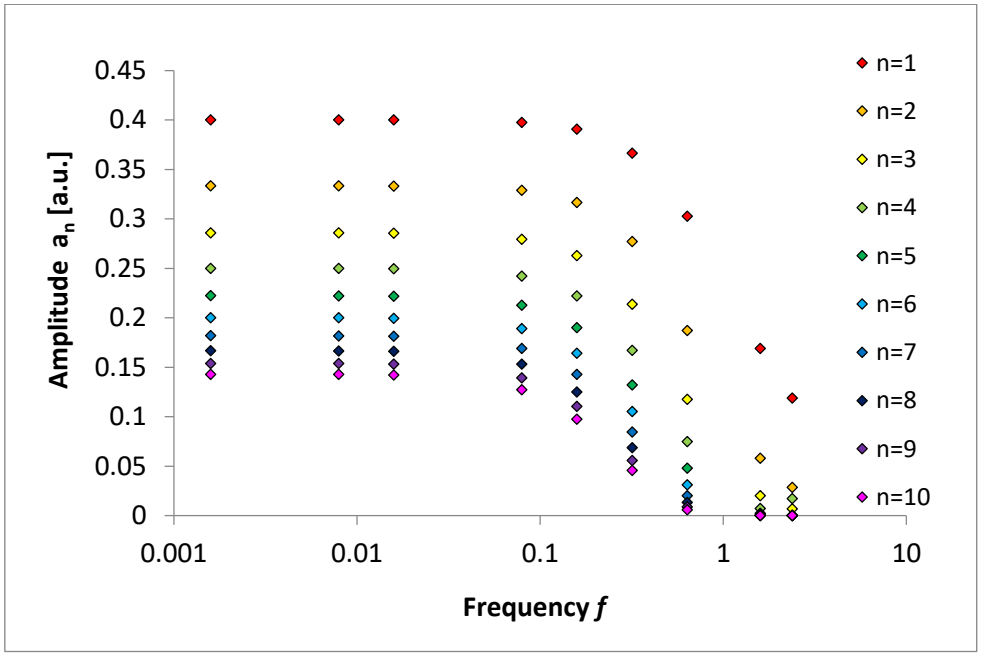


Figure 93. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f =1 *u*/s^{-1}, C =4 *u*, k_r =4 s^{-1}, A_0=0.5 *u*, B_0=0.5 *u*, where *u* is a unit of concentration.

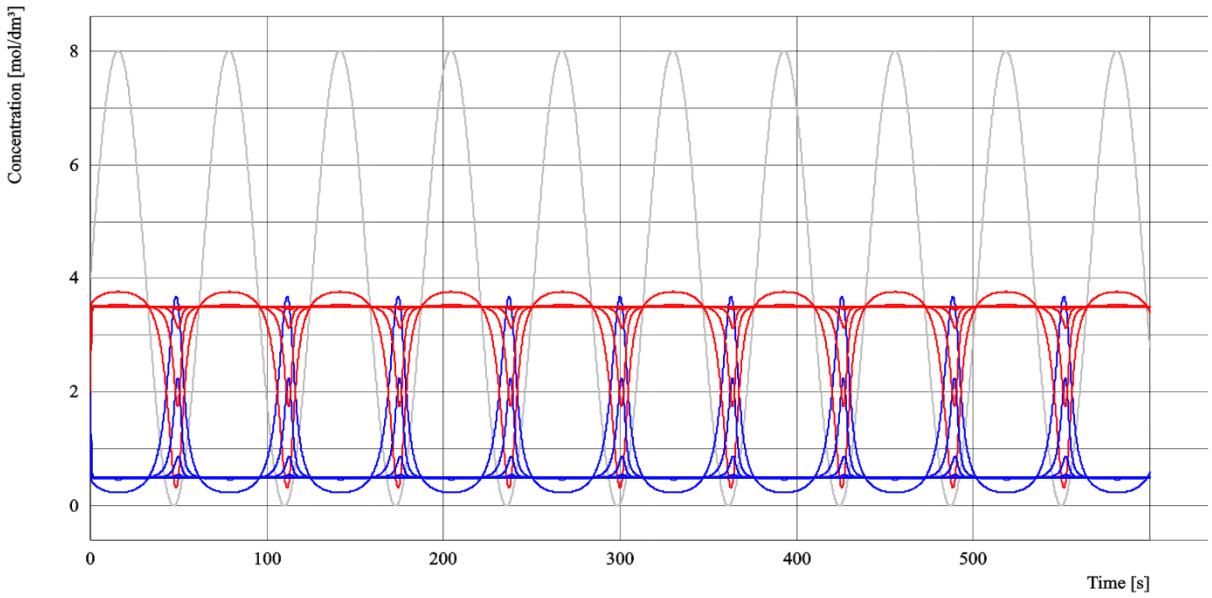


Figure 94. The concentration profile of a catalytic cascade. k_f=1 *u*/s^{-1}, C=4 *u*, k_r=0.5 s^{-1}, A_0=4 *u*, B_0=4 *u*, where *u* is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 3.5\ u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

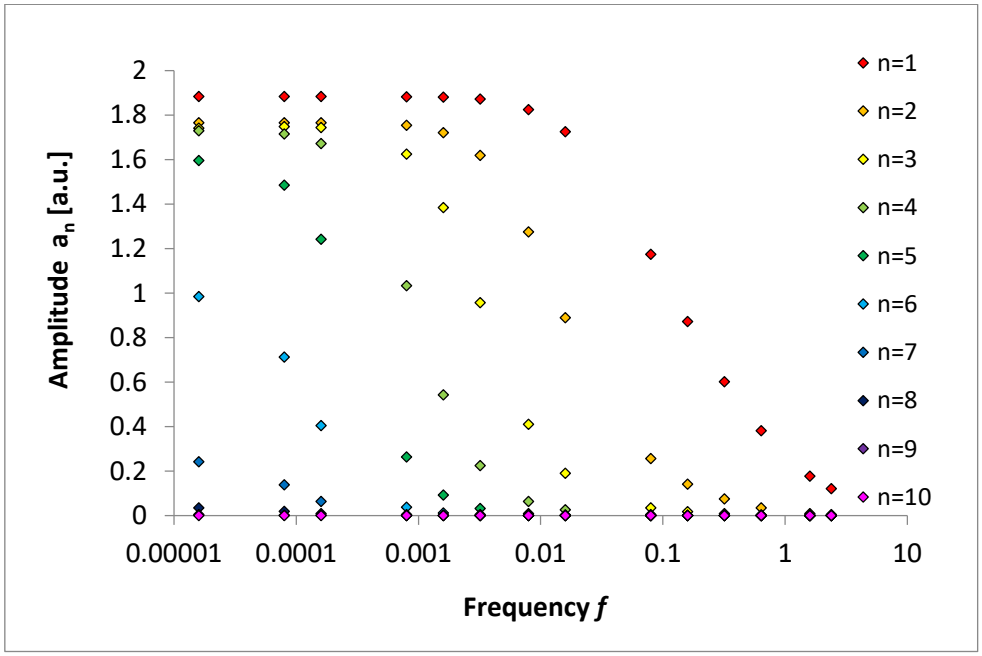


Figure 95. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f=1 *u*/s^{-1}, C=4 *u*, k_r=0.5 s^{-1}, A_0=4 *u*, B_0=4 *u*, where *u* is a unit of concentration.

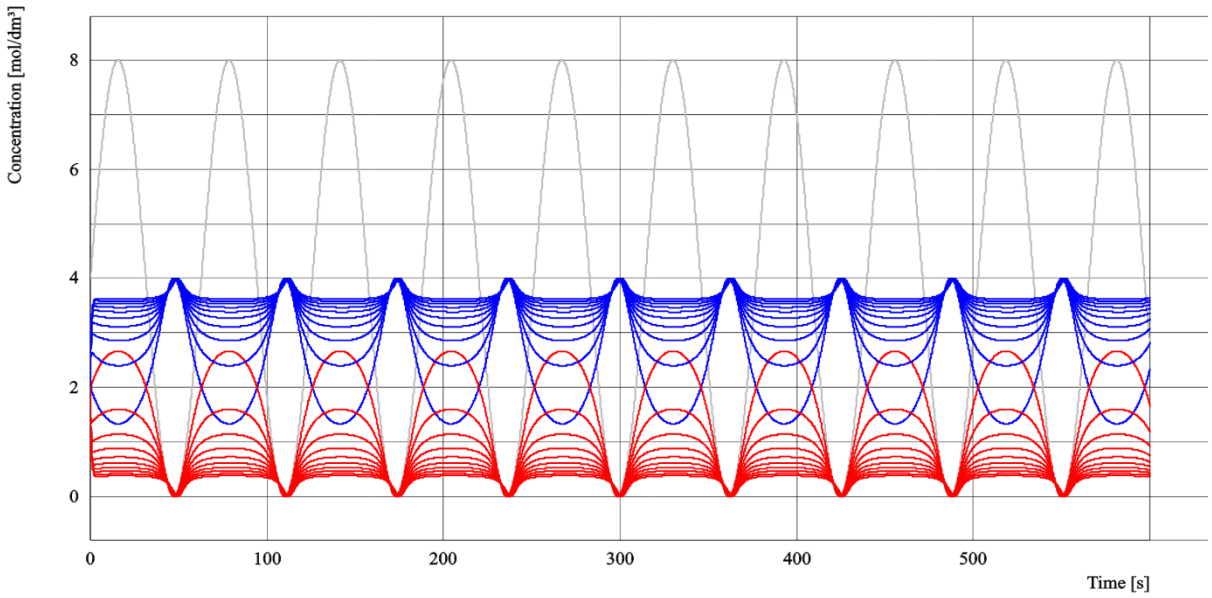


Figure 96. The concentration profile of a catalytic cascade. k_f**=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r**=4 s^{-1}**, A_0**=4** ***u*****,** B_0**=4** ***u***, where *u* is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 0$ *u* with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

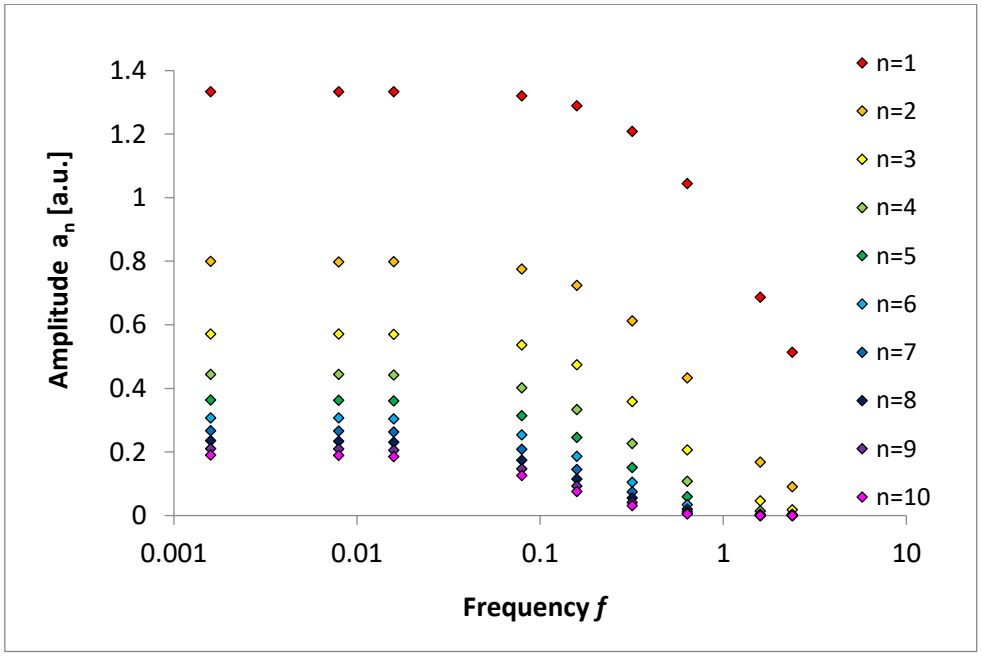


Figure 97. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f**=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r**=4 s^{-1}**, A_0**=4** ***u*****,** B_0**=4** ***u***, where *u* is a unit of concentration.

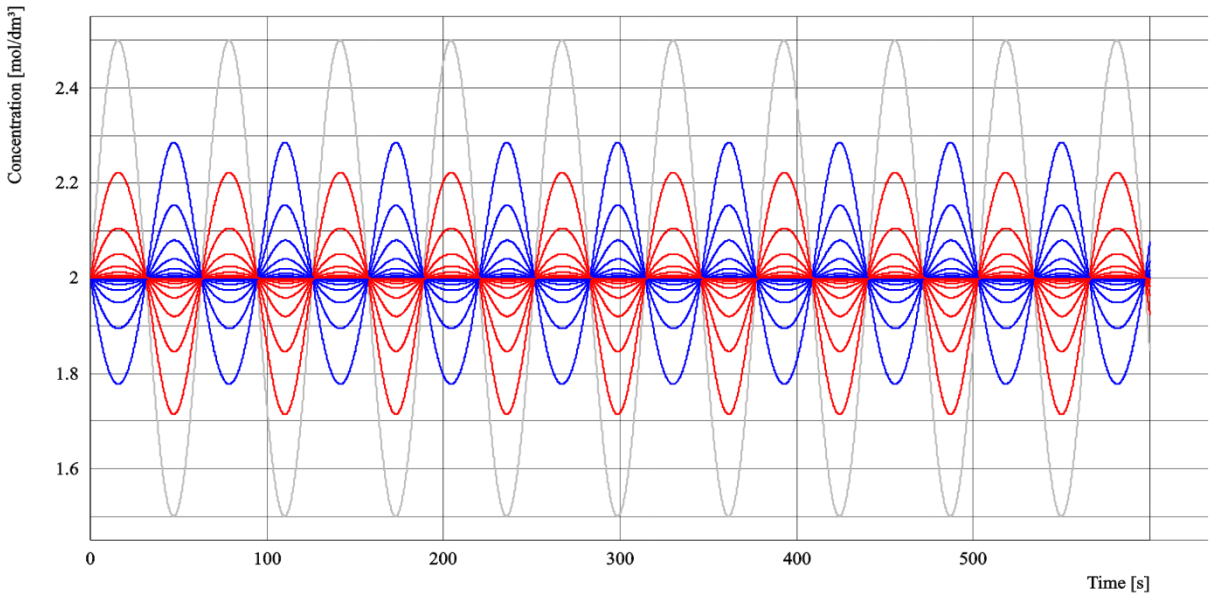


Figure 98. The concentration profile of a catalytic cascade. k_f=**2** u/s^{-1}, C =**4** u, k_r=**4** s^{-1}, A_0 =**4** u, B_0 =**2** u, where u is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 2\,u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

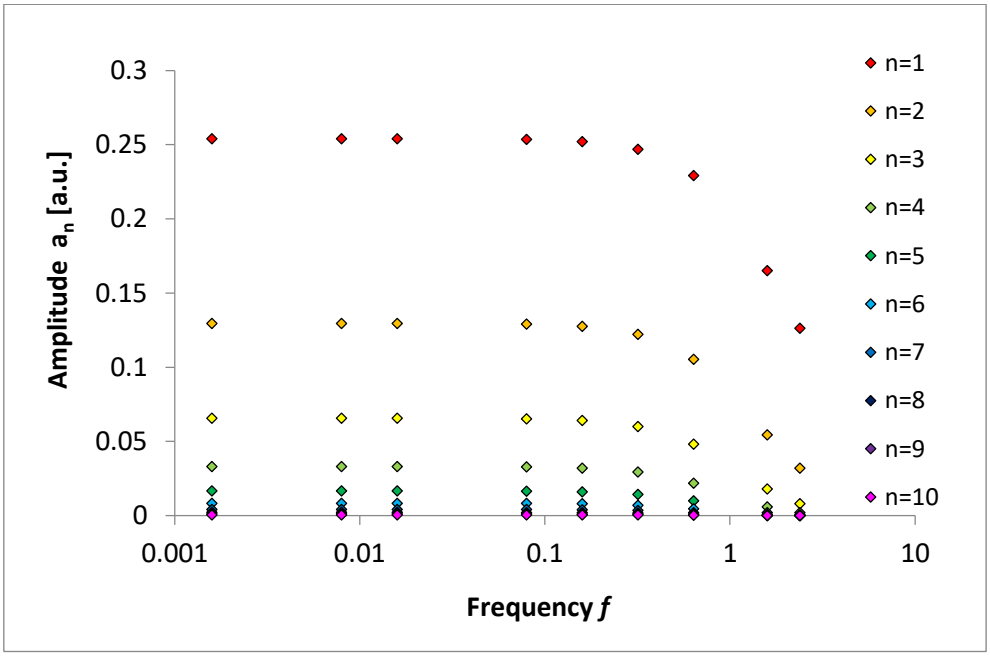


Figure 99. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the n-numbered active form. k_f=**2** u/s^{-1}, C =**4** u, k_r=**4** s^{-1}, A_0 =**4** u, B_0 =**2** u, where u is a unit of concentration.

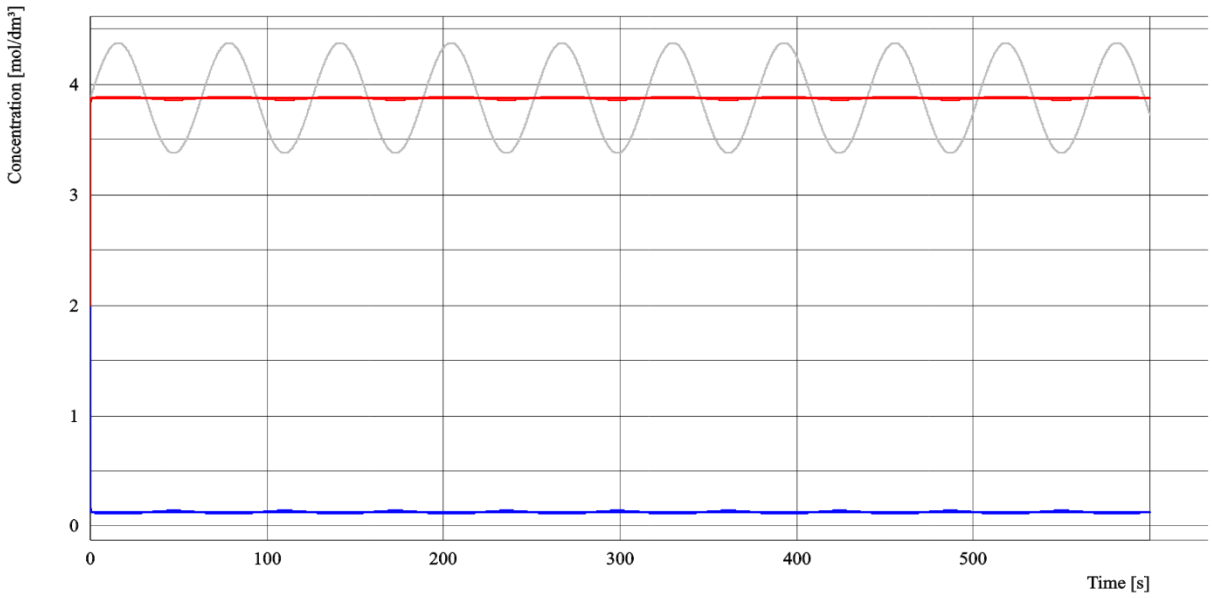


Figure 100. The concentration profile of a catalytic cascade. k_f **=4** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=0.5 s^{-1}**, A_0**=0.5** ***u*****,** B_0 **=3.875** ***u***, where *u* is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 3.875\ u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

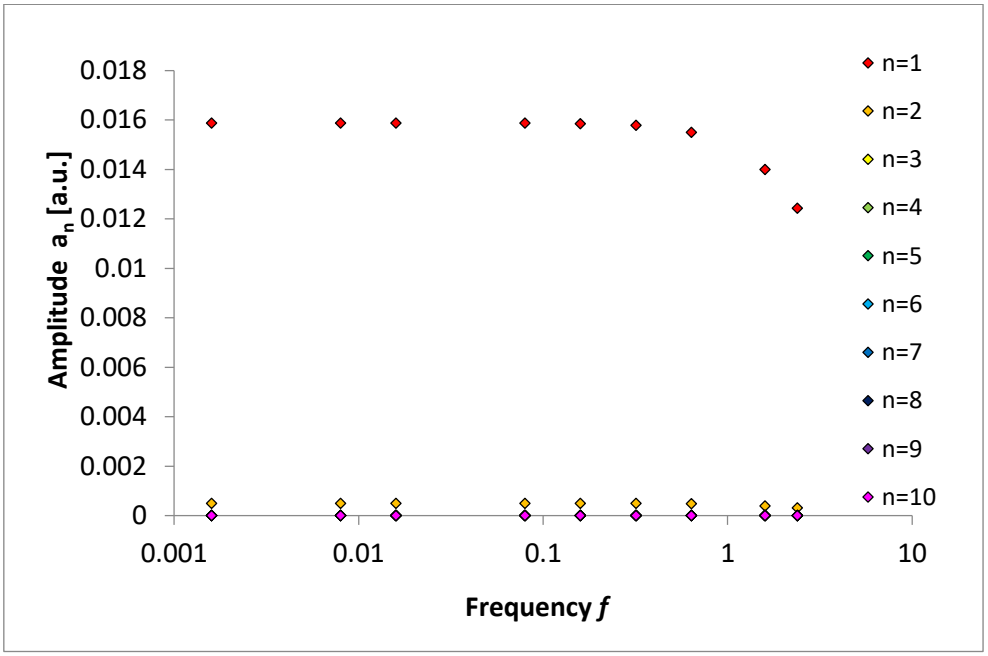


Figure 101. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f **=4** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=0.5 s^{-1}**, A_0**=0.5** ***u*****,** B_0 **=3.875** ***u***, where *u* is a unit of concentration.

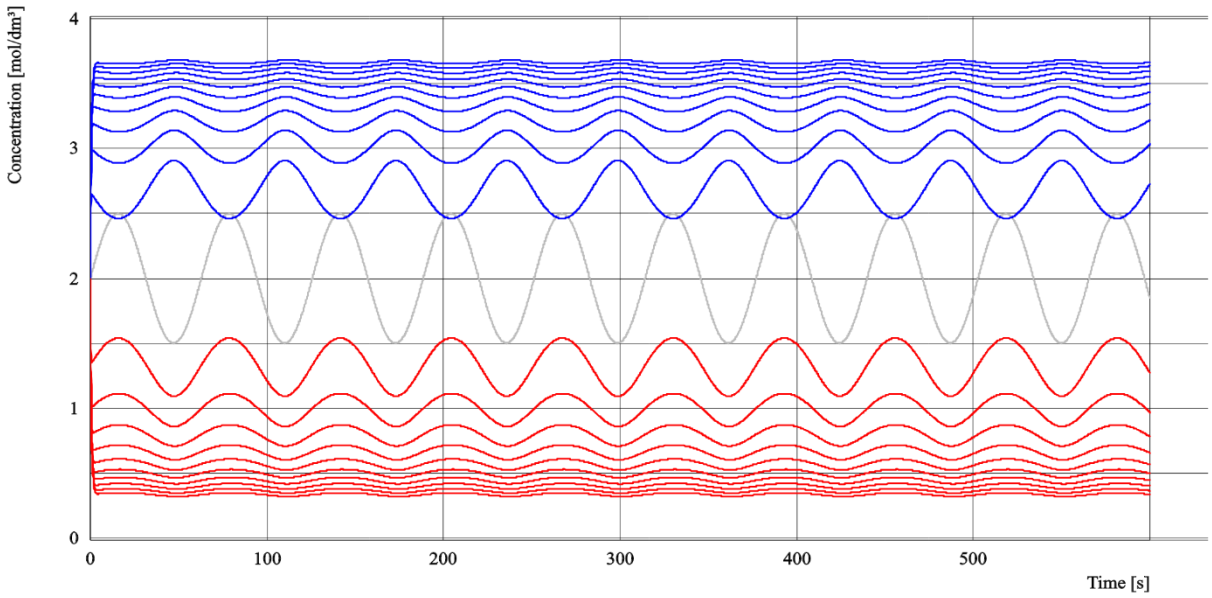


Figure 102. The concentration profile of a catalytic cascade. k_f **=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=4 s^{-1}**, A_0**=0.5** ***u*****,** B_0 **=2** ***u***, where *u* is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 0$ *u* with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

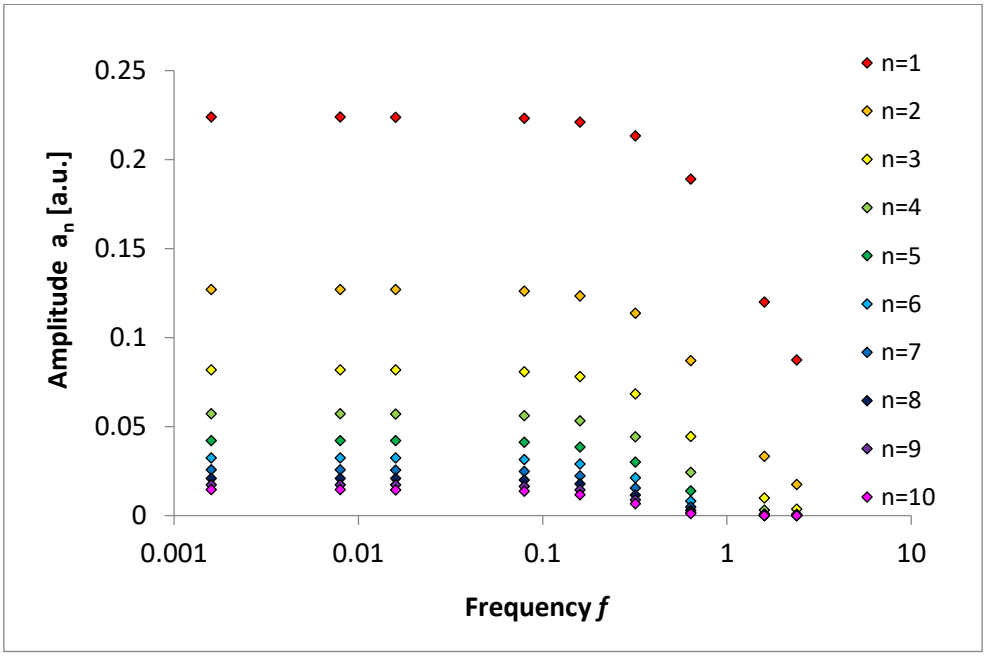


Figure 103. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f **=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=4 s^{-1}**, A_0**=0.5** ***u*****,** B_0 **=2** ***u***, where *u* is a unit of concentration.

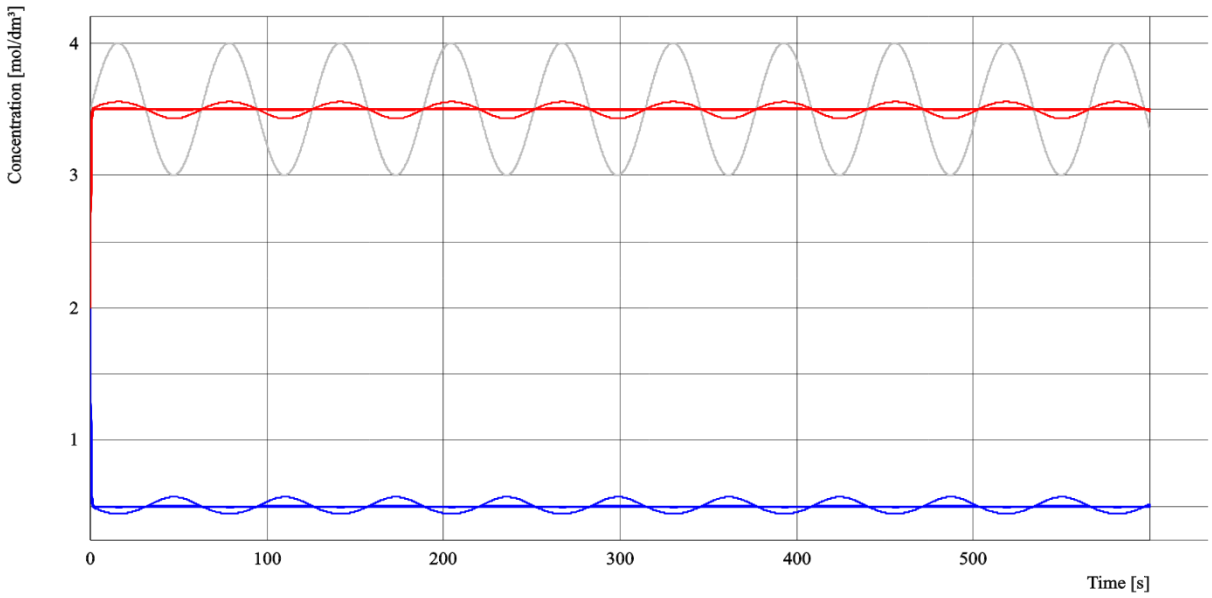


Figure 104. The concentration profile of a catalytic cascade. k_f **=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=0.5 s^{-1}**, A_0**=0.5** ***u*****,** B_0**=3.5** ***u***, where *u* is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 3.5\ u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

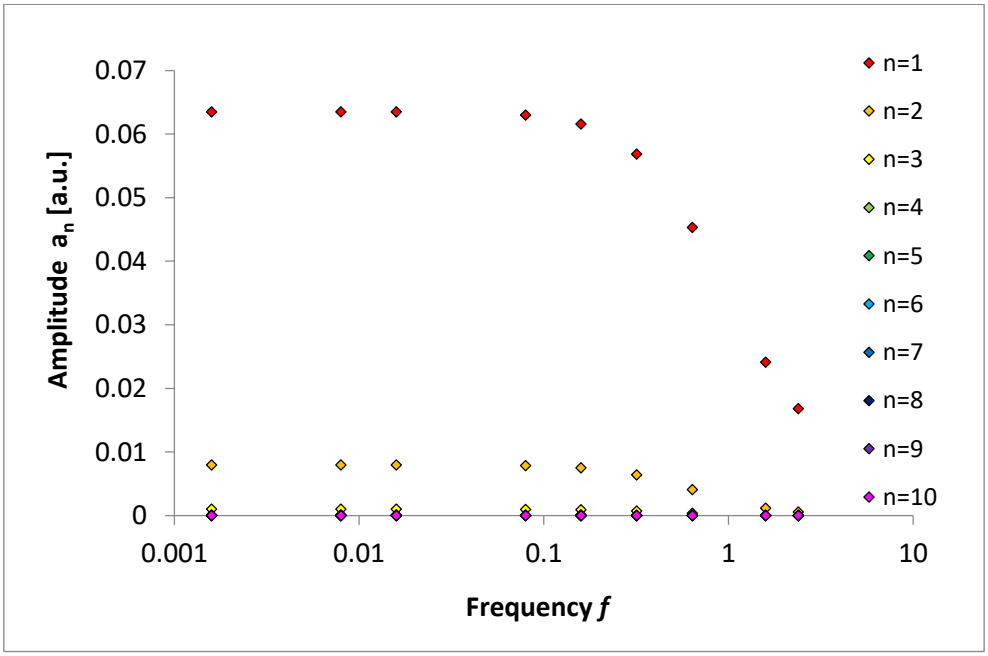


Figure 105. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f **=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=0.5 s^{-1}**, A_0**=0.5** ***u*****,** B_0**=3.5** ***u***, where *u* is a unit of concentration.

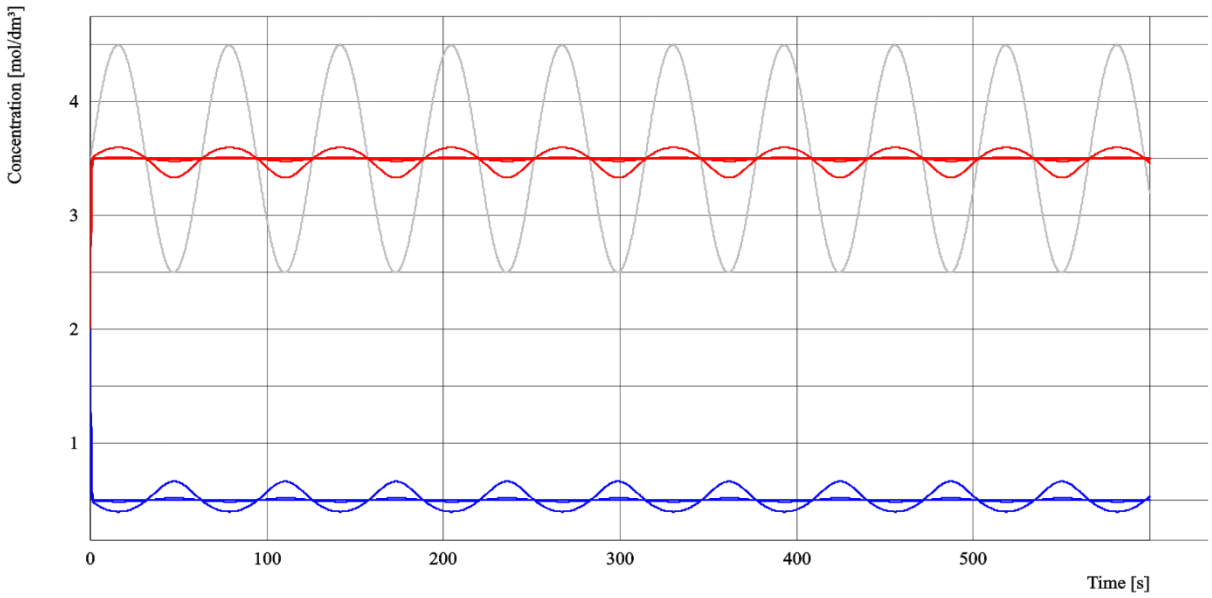


Figure 106. The concentration profile of a catalytic cascade. k_f **=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=0.5 s^{-1}**, A_0**=1** ***u*****,** B_0 **=3.5** ***u***, where *u* is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 3.5\ u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

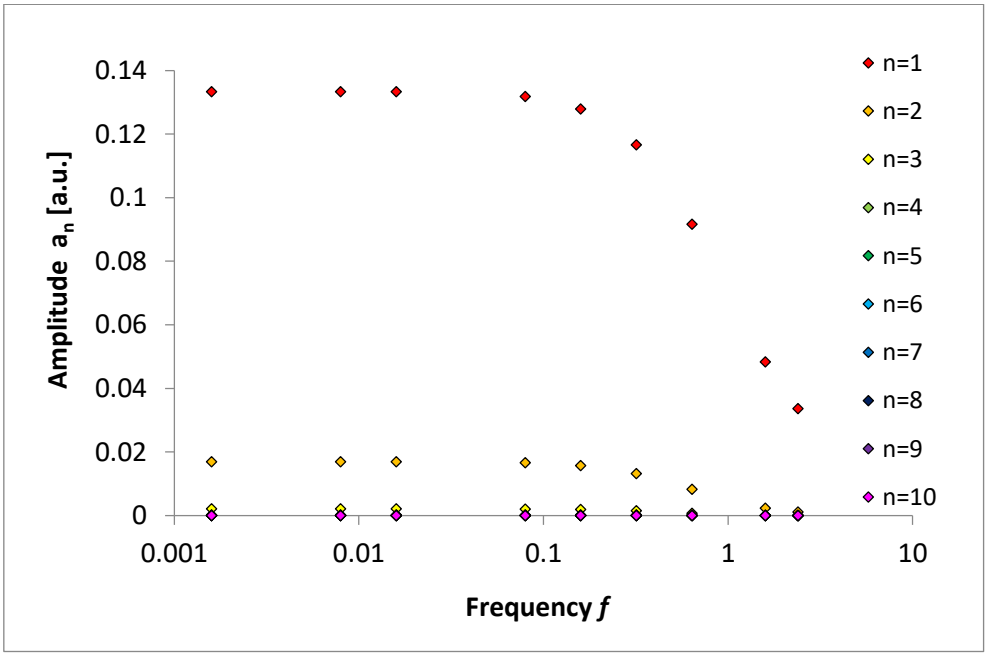


Figure 107. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f **=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=0.5 s^{-1}**, A_0**=1** ***u*****,** B_0 **=3.5** ***u***, where *u* is a unit of concentration.

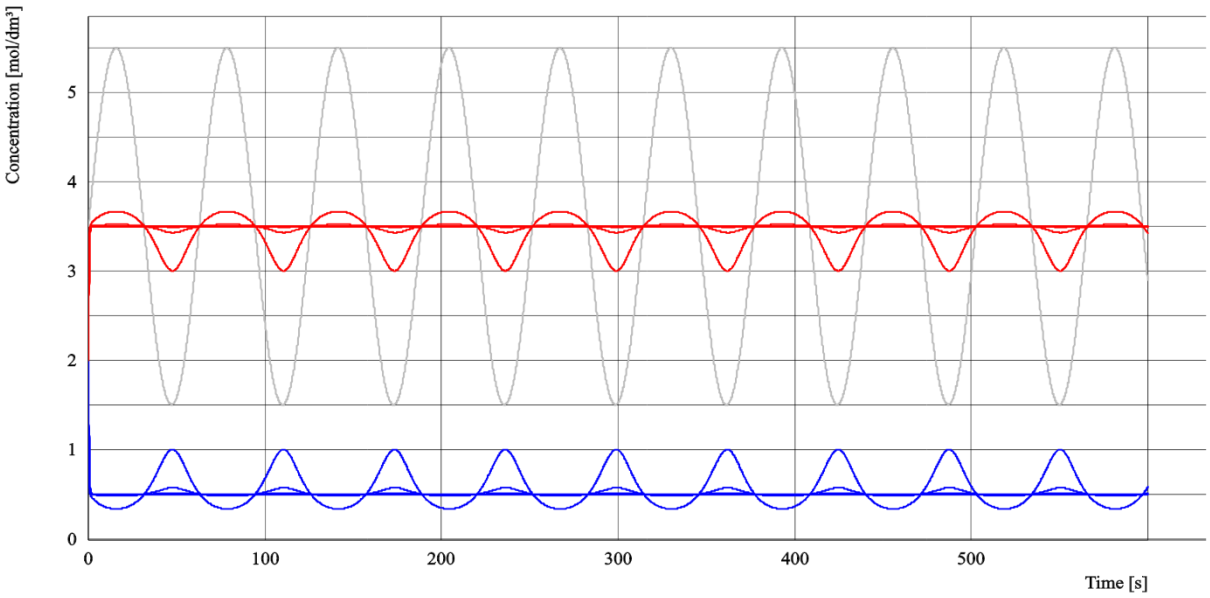


Figure 108. The concentration profile of a catalytic cascade. k_f **=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=0.5 s^{-1}**, A_0**=2** ***u*****,** B_0 **=3.5** ***u***, where *u* is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 3.5\ u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

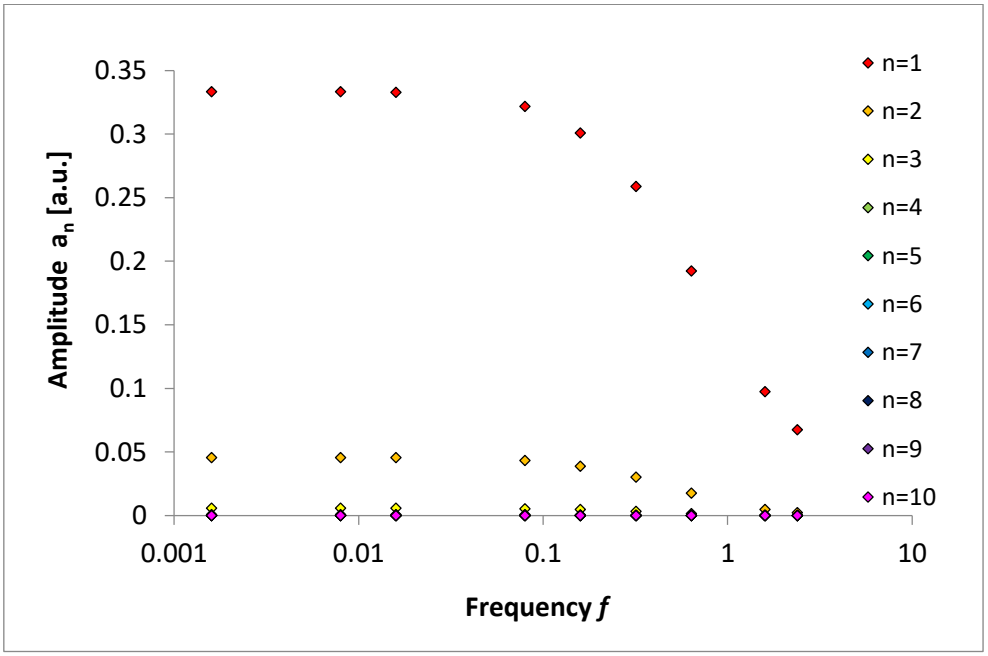


Figure 109. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of *n*-numbered active form. k_f **=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=0.5 s^{-1}**, A_0 **=2** ***u*****,** B_0 **=3.5** ***u***, where *u* is a unit of concentration.

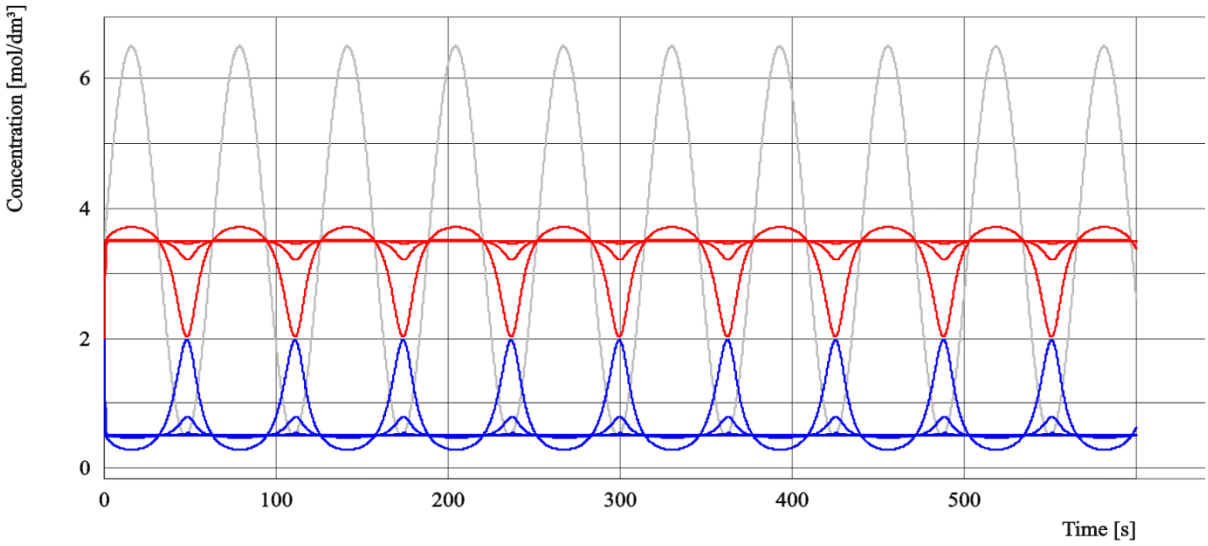


Figure 110. The concentration profile of a catalytic cascade. k_f **=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=0.5 s^{-1}**, A_0**=3** ***u*****,** B_0 **=3.5** ***u***, where *u* is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 3.5\ u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

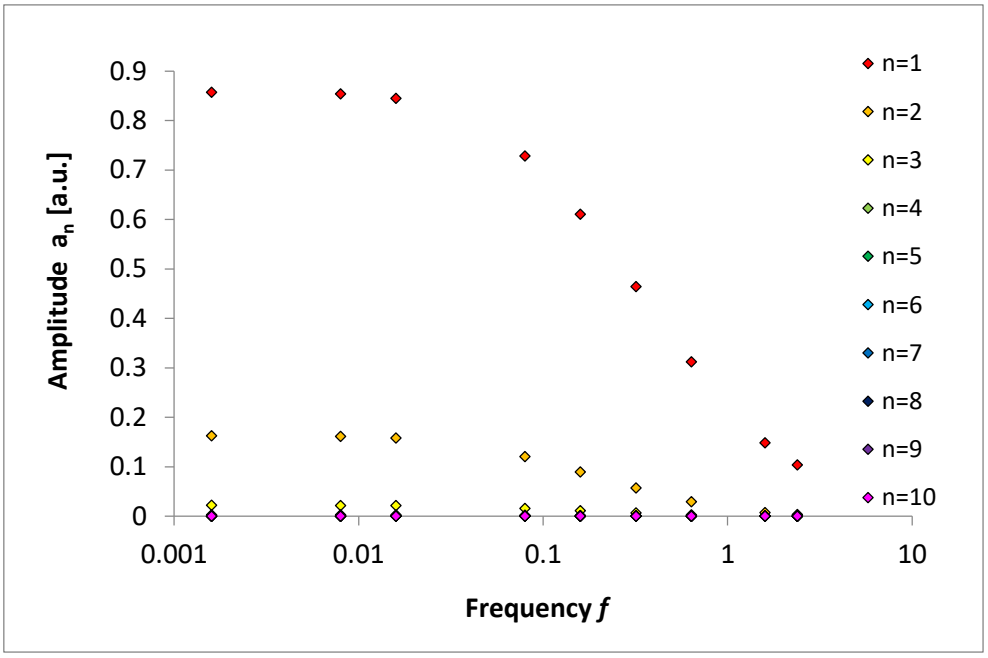


Figure 111. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f **=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=0.5 s^{-1}**, A_0**=3** ***u*****,** B_0 **=3.5** ***u***, where *u* is a unit of concentration.

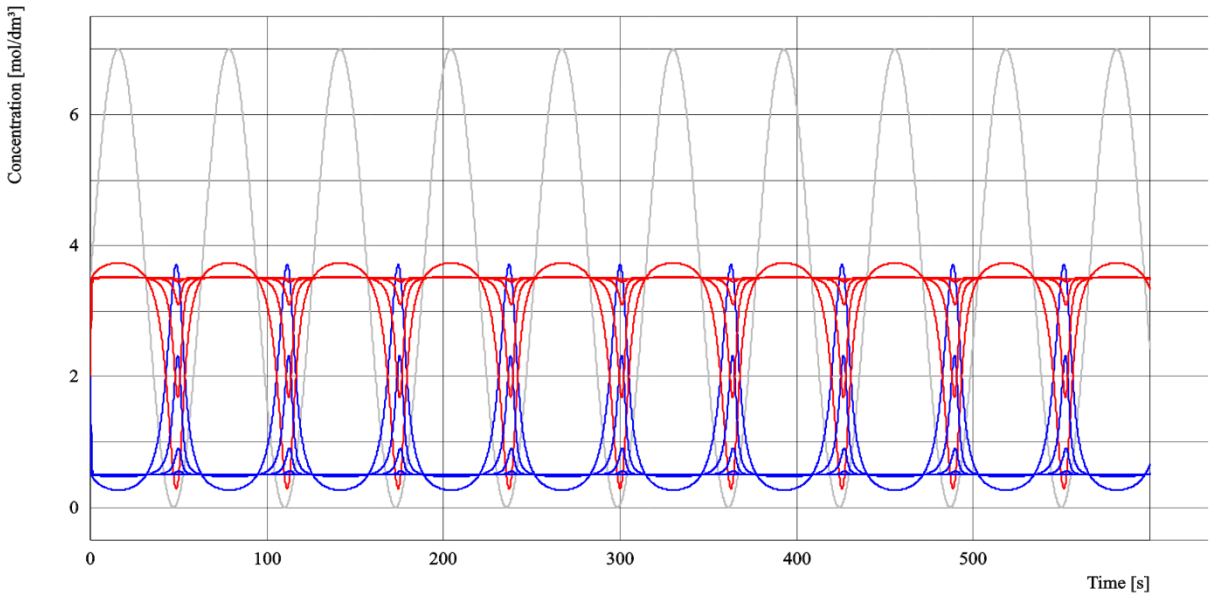


Figure 112. The concentration profile of a catalytic cascade. k_f **=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=0.5 s^{-1}**, A_0 **=3.5** ***u*****,** B_0 **=3.5** ***u***, where *u* is a unit of concentration. The inactive form apparently tends to the concentration of $B - \frac{k_r}{k_f} = 3.5\ u$ with each consecutive layer. The grey line denotes the input signal, blue lines – inactive forms, red lines – active forms.

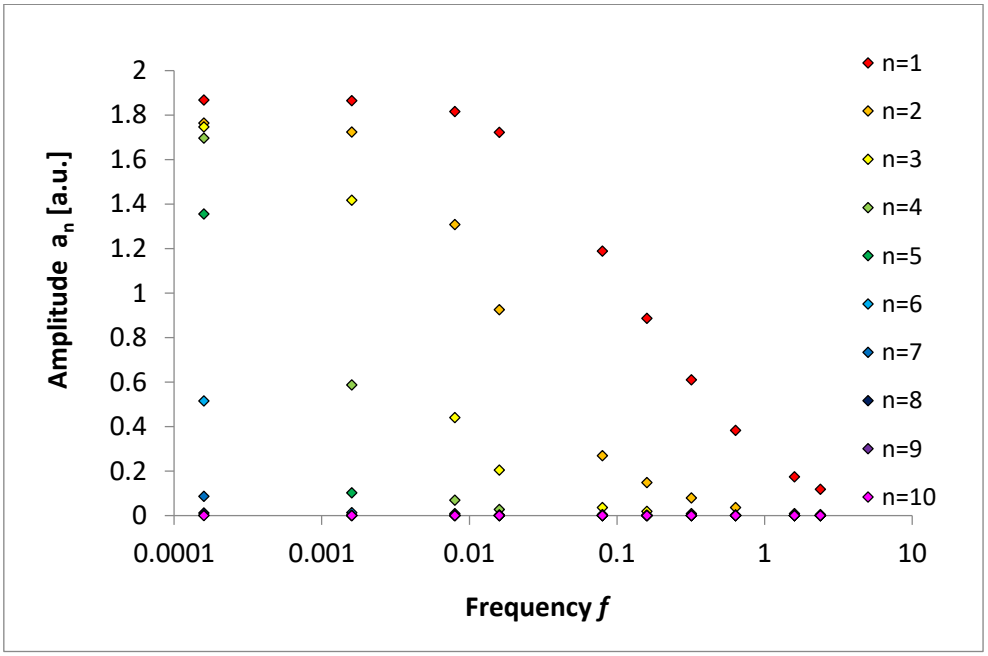


Figure 113. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f **=1** ***u*****/s^{-1}**, ***C*** **=4** ***u***, k_r **=0.5 s^{-1}**, A_0**=3.5** ***u*****,** B_0 **=3.5** ***u***, where *u* is a unit of concentration.

The plots presented in **Figures 82-113** were largely inconsistent in respect of their shapes – it was deduced that, unlike their linear predecessors, the cascades exhibit two or more "working modes" – which mode applies in a given case depends on the network parameters.

Assuming there were two "modes", the cases shown in **Figures 84-89, 94-95, 98-99** and **112-113** represent the same mode, as their amplitude-frequency plots were significantly different from the rest. Their common characteristic was that in their time-concentration profiles, the absolute concentrations of some active forms a_n crossed the baseline of the inactive forms $B_n' = B_n - \frac{k_{rev}}{k_{fr}}$ at some points – on the other hand, the remaining plots did not show such crossing (further explanation is available in **Figure 114**).

It took another 100+ simulations to establish the expression for a "switch" between the modes, which was supposed to enable the further *a priori* prediction of k_{eff}. According to this expression, the concentration profiles do not cross the baselines (the first mode) unless:

$$\frac{k_{rev}}{k_{fw}C} \geq 0.6192 \left(\frac{A_0 + B_0}{C}\right)^{0.218151},$$

otherwise the "curve crossing" is observed (the second mode). Albeit the equation itself allows for predicting whether the "crossing" occurs with an arguable precision, it proved to be an unreliable indicator of which "mode" is in effect, if not at extremes. Even worse, this equation only concerned the homogeneous pathways, thus another approach was necessary in order to find the effective rate constant k_{eff}.

There were, indeed, two "modes" of operation – one, where all oscillations were very strongly dampened, but the threshold frequency was more apparent, and another, with a stronger a long-range transmission, but the elusive threshold frequency due to the significant frequency "shifts" between layers *n*.

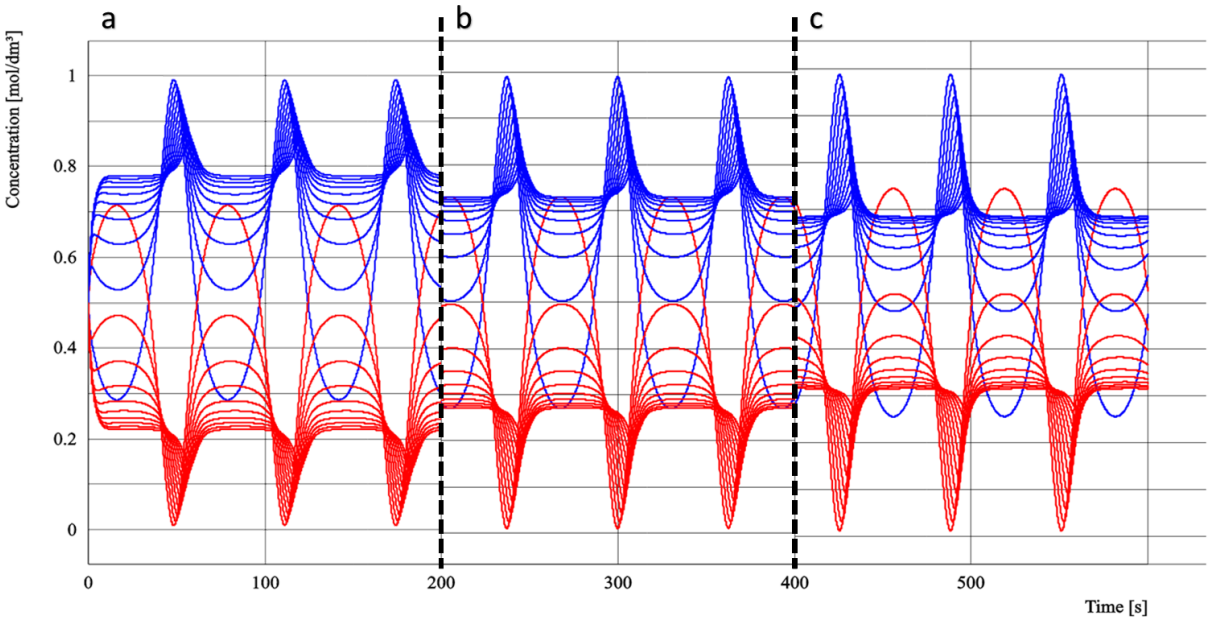


Figure 114. Each segment of the plot shows the concentration profile of a homogeneous catalytic cascade with different reverse rate constants. In all segments (separated by the black dotted lines) k_f**=1** ***u*****/s^{-1}**, ***C*** **=1** ***u***, A_0 **=1** ***u*****,** B_0 **=1** ***u***, where *u* is an arbitrary unit of concentration. The reverse rate constant k_r equals 0.8 s^{-1} in segment (a), 0.74 s^{-1} in segment (b), and 0.69 s^{-1} in segment (c). The blue lines denote the inactive forms, whereas red lines denote the active forms. Notice that the concentration of the active form in (a) is always lower than $\frac{k_{rev}}{k_{fr}}$, but in (b) the concentration of the first active form $n = 1$ periodically becomes equal to $\frac{k_{rev}}{k_{fr}}$, which is observed as the tallest red curve "tangentially touching" the uppermost blue curve – this was considered an indicator of a transition state between the two modes of operation. The further lowering of k_r, as seen in (c), resulted in the concentration of the first active form periodically exceeding the $\frac{k_{rev}}{k_{fr}}$ value. Thus, the concentration profile (a) was supposed to indicate one mode of operation, (c) another mode and (b) the transition between these two. The subsequent experiments, shown in section **2.6.2**, proved that this method was unreliable.

According to the apparent significance of the $\frac{k_{rev}}{k_{fw}}$ expression, it was considered as a single variable, effectively reducing the problem from 3*n*+1 to 2*n*+1 variables, i.e., $\frac{k_n^{rev}}{k_n^{fw}}$, C_n and B_0. In the cases when the cascades were homogeneous:

$$k_1^{fw} = k_2^{fw} = \cdots = k_{fw}, \qquad k_1^{rev} = k_2^{rev} = \cdots = k_{rev}, \qquad C_1 = C_2 = \cdots = C,$$

there were only 3 variables left – $\frac{k_{rev}}{k_{fw}}$, C and B_0.

A series of simulations have been performed, with varying $\frac{k_{rev}}{k_{fw}}$ and other variables constant at 1 a.u. – the resulting plots, shown in **Figures 115-119**, allowed for the discovery that at particular frequencies, the signal strength at the (*n*+1)th step is either $\frac{k_{rev}}{k_{fw}}$ or $\left(\frac{k_{rev}}{k_{fw}}\right)^{-1}$ times weaker than at the n^{th} step.

As shown in **Table 4**, when $\frac{k_{rev}}{k_{fw}} > 1\ a.u.$ and the signal frequency belonged to the transmissive domain, a $\frac{k_{rev}}{k_{fw}}$-fold amplitude decrease was observed at each consecutive layer. However, in the case when $\frac{k_{rev}}{k_{fw}} < 1\ a.u.$ and the signal frequency belonged to the threshold domain, a $\left(\frac{k_{rev}}{k_{fw}}\right)^{-1}$-fold amplitude decrease was observed.

k_{rev}/k_{fw}	A_0/A_1	A_1/A_2	A_2/A_3	A_3/A_4	A_4/A_5	A_5/A_6	A_6/A_7	A_7/A_8	A_8/A_9	A_9/A_{10}
0.5	2.501	1.303	1.122	1.067	1.059	1.085	1.152	1.268	1.433	1.616
1	3.000	1.667	1.400	1.286	1.223	1.182	1.155	1.134	1.119	1.106
2	4.000	2.500	2.200	2.091	2.044	2.021	2.011	2.005	2.003	2.001
4	6.000	4.333	4.077	4.019	4.005	4.001	4.000	4.000	4.000	4.000
10	12.000	10.167	10.016	10.002	10.000	10.000	10.000	10.000	10.000	10.000
0.5*	2.593	1.460	1.407	1.525	1.694	1.846	1.950	2.008	2.036	2.048

Table 4. The amplitude ratios from the plots shown in **Figures 115-119**, observed in the transmissive domains at ω=0.01 s^{-1}, f=0.0016 s^{-1} and in the threshold domain at ω=0.1 s^{-1}, f=0.016 s^{-1}, with the latter marked with an asterisk. The remaining variables were set at 1 a.u.

While the observations made on the basis of **Figures 115-119** were too specific to have any practical applications, they helped establish the $\frac{k_{rev}}{k_{fw}}$ value as an important determinant of both the "mode" by which the cascade works and its effective rate constant k_{eff}. The subsequent simulations, shown in **Figures 120-121**, allowed for verifying that changing the absolute values of k_{fw} and k_{rev} had no effect as long as $\frac{k_{rev}}{k_{fw}}$ remained the same.

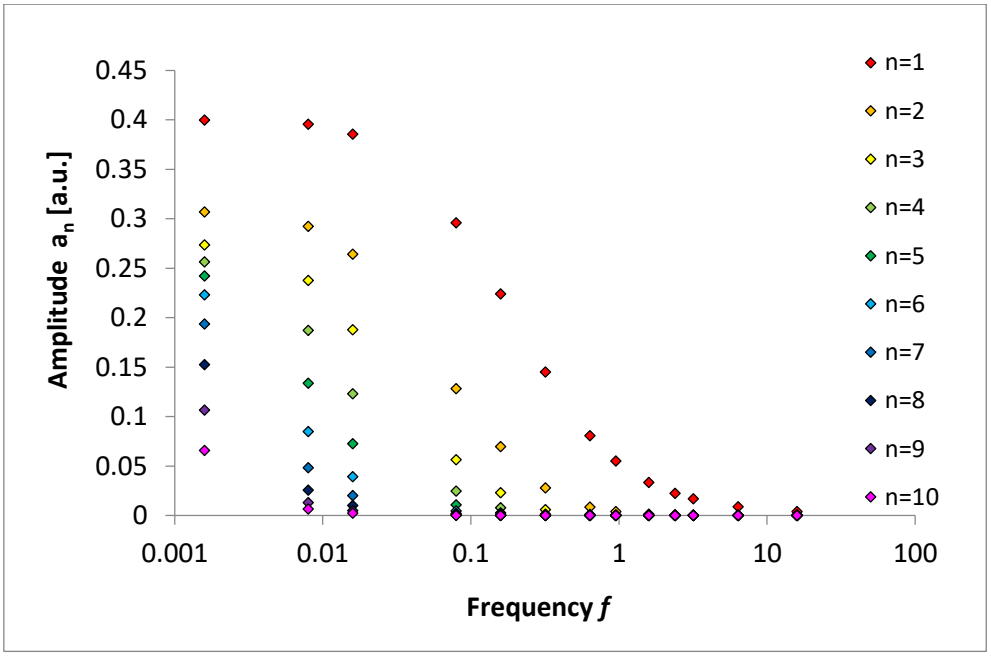


Figure 115. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f **=1** u/s^{-1}, C **=1** u, k_r **=0.5** s^{-1}, A_0**=1** u, B_0 **=1** u, where *u* is an arbitrary unit of concentration. $\frac{k_r}{k_f} = 0.5\ u$.

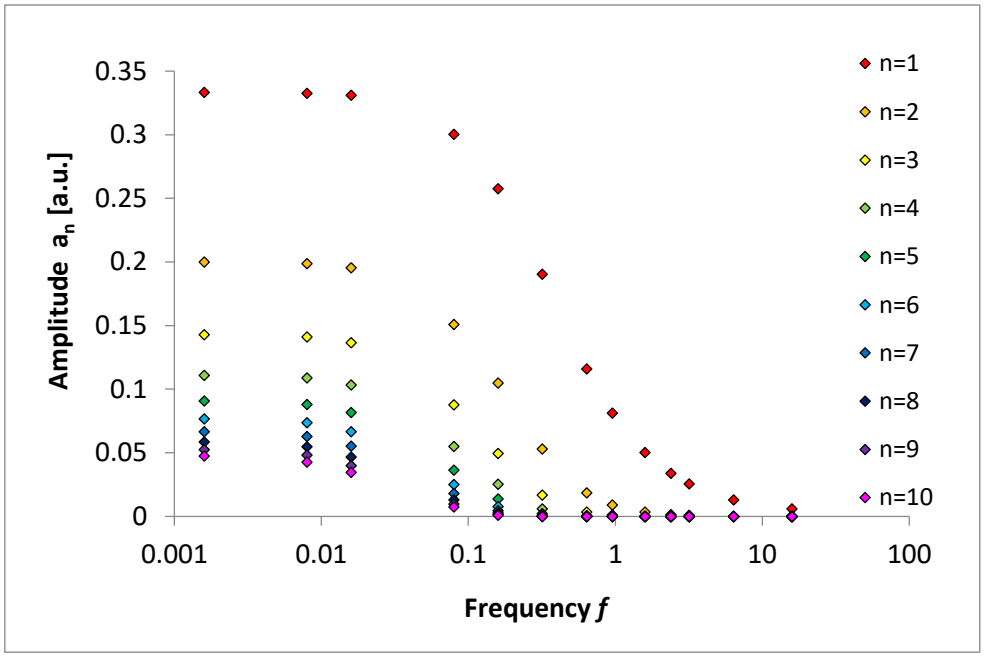


Figure 116. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f**=1** u/s^{-1}, C**=1** u, k_r**=1** s^{-1}, A_0**=1** u**,** B_0**=1** u, where u is an arbitrary unit of concentration. $\frac{k_r}{k_f} = 1\ u$.

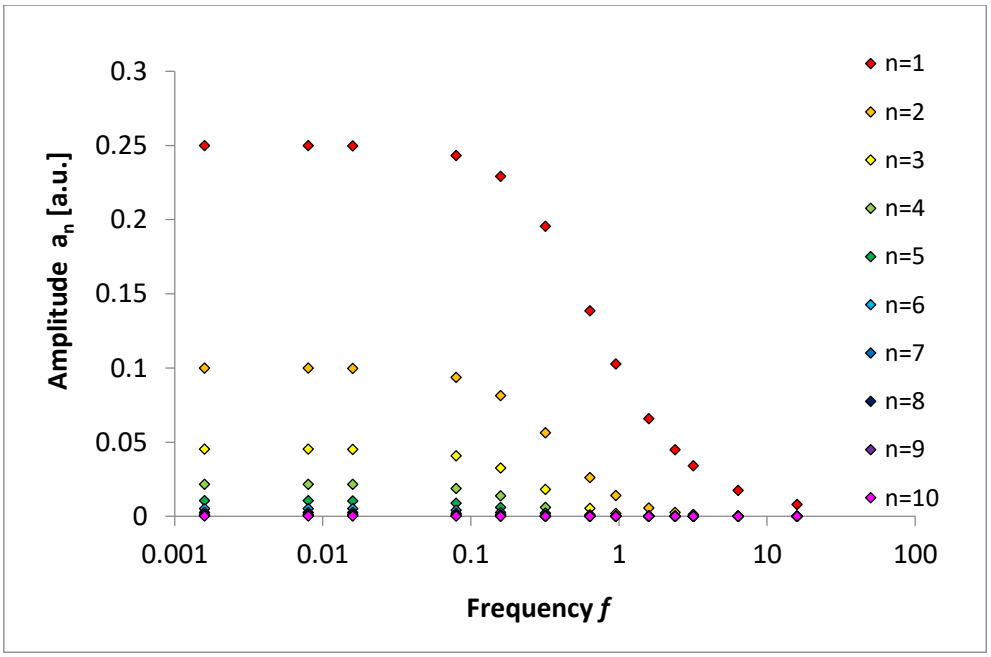


Figure 117. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f**=1** u/s^{-1}, C**=1** u, k_r**=2** s^{-1}, A_0**=1** u**,** B_0**=1** u, where u is an arbitrary unit of concentration. $\frac{k_r}{k_f} = 2\ u$.

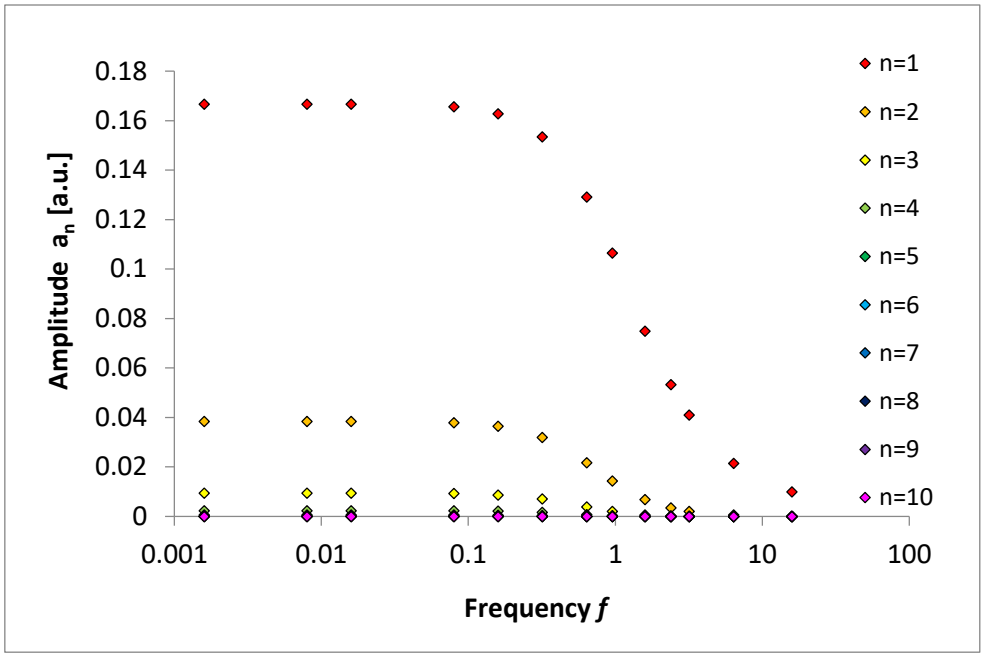


Figure 118. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f**=1** u/s^{-1}, C**=1** u, k_r**=4** s^{-1}, A_0**=1** u**,** B_0**=1** u, where u is an arbitrary unit of concentration. $\frac{k_r}{k_f} = 4\ u$.

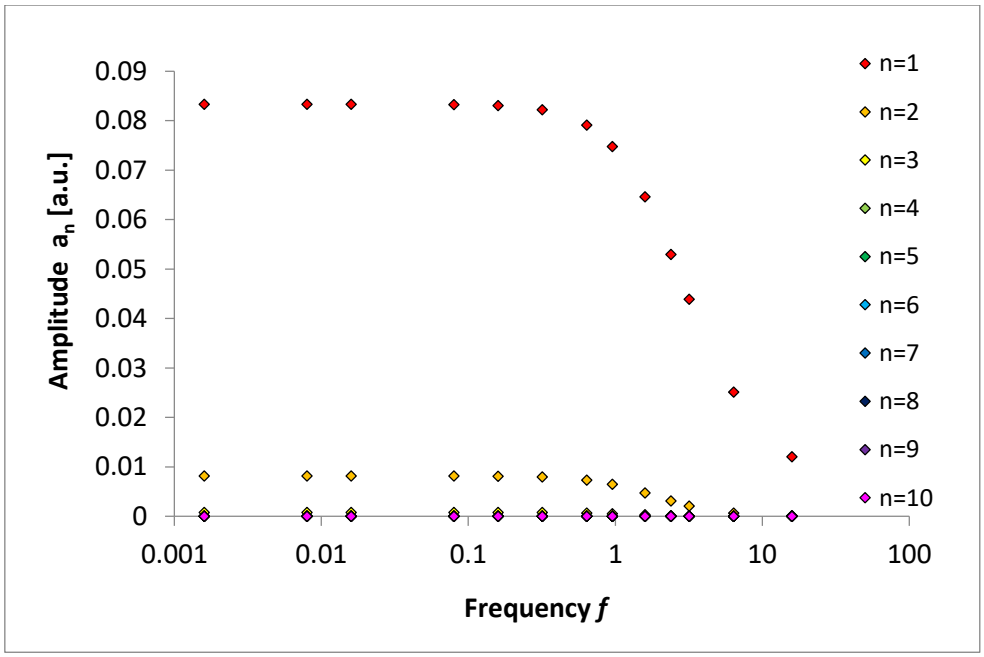


Figure 119. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f**=1** u/s^{-1}, C**=1** u, k_r**=10** s^{-1}, A_0**=1** u**,** B_0**=1** u, where u is an arbitrary unit of concentration. $\frac{k_r}{k_f} = 10\ u$.

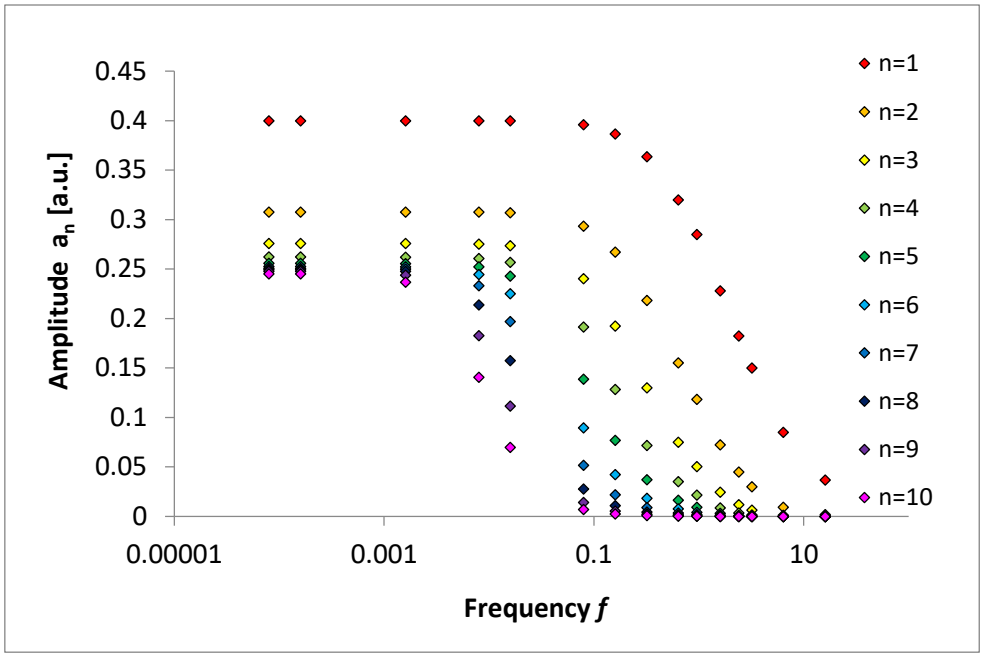


Figure 120. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f**=10** u/s^{-1}, C**=1** u, k_r**=5** s^{-1}, A_0**=1** u**,** B_0**=1** u, where *u* is an arbitrary unit of concentration. $\frac{k_r}{k_f} = 0.5\ u$.

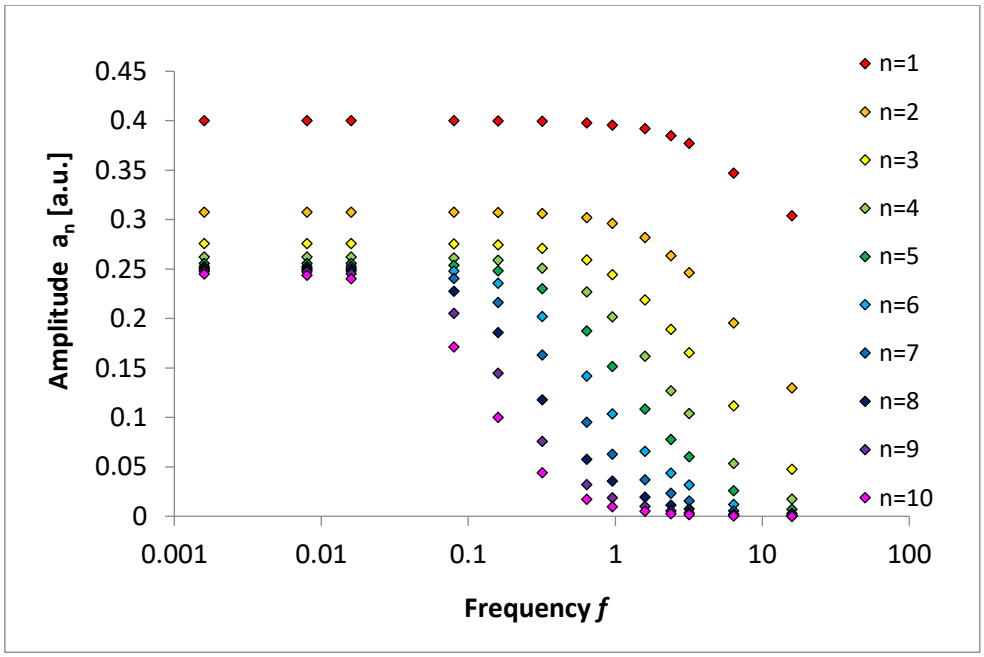


Figure 121. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f**=100** u/s^{-1}, C**=1** u, k_r**=50** s^{-1}, A_0**=1***u*, B_0**=1** u, where *u* is an arbitrary unit of concentration. $\frac{k_r}{k_f} = 0.5\ u$.

The further analysis of the plots presented in **Figures 120-122** shows how the "working mode" of the cascade affects its transmissive properties. As seen in all the previous plots, the signals with the frequencies belonging to the transmissive domain were significantly weaker after many cascade layers – contrary to the linear pathways, where they would retain almost full strength. This feature was investigated in **Table 5** – if $\frac{k_{rev}}{k_{fw}} > 1\ a.u.$, then the amplitude of a signal with the frequency belonging to the transmissive domain decreased $\frac{k_{rev}}{k_{fw}}$ per step times at distant layers – falling to almost zero in the process. However, if $\frac{k_{rev}}{k_{fw}} < 1\ a.u.$, then the asymptotic behavior was observed, as the amplitude eventually stopped decreasing (i.e., $A_n/A_{n+1} \approx 1$).

k_{rev}/k_{fw}	A_0/A_1	A_1/A_2	A_2/A_3	A_3/A_4	A_4/A_5	A_5/A_6	A_6/A_7	A_7/A_8	A_8/A_9	A_9/A_{10}
5/10	2.500	1.300	1.115	1.052	1.025	1.013	1.007	1.006	1.007	1.011
50/100	2.500	1.300	1.115	1.052	1.025	1.013	1.007	1.006	1.007	1.011
10/5	4.000	2.500	2.200	2.091	2.043	2.021	2.011	2.005	2.003	2.001
5/10*	2.586	1.448	1.388	1.499	1.667	1.823	1.932	1.996	2.028	2.042

Table 5. The amplitude ratios from the plots shown in **Figures 120-122**, observed in the transmissive domains (ω=0.0005 s^{-1}, f=0.0008 s^{-1} for **5/10**; ω=0.01 s^{-1}, f=0.0016 s^{-1} for **100/50** and **10/5**) or in the threshold domain (ω=1 s^{-1}, f=0.1592 s^{-1}) – the latter is marked with an asterisk. The remaining variables were set at 1 a.u.

While changing the absolute values of k_{fw} and k_{rev} had no significant effect on the signal as long as $\frac{k_{rev}}{k_{fw}}$ remained constant, the same was not true for *C*, which was thus far set constant at 1 a.u. – changing the value of catalyst total concentration on every layer could potentially "switch the mode". As shown in **Figures 123-126** and **Table 6**, it was actually the dimensionless $\frac{k_{rev}}{k_{fw}C}$ ratio that determined transmissive properties of the cascade – for $\frac{k_{rev}}{k_{fw}C} > 1$, the signals with low frequencies decreased $\frac{k_{rev}}{k_{fw}C}$ –fold per step at distant layers, while for $\frac{k_{rev}}{k_{fw}C} < 1$, the asymptotic behavior was observed.

k_{rev}/k_{fw}	$k_{rev}/k_{fw}C$	A_0/A_1	A_1/A_2	A_2/A_3	A_3/A_4	A_4/A_5	A_5/A_6	A_6/A_7	A_7/A_8	A_8/A_9	A_9/A_{10}
0.5	**0.25**	1.251	1.054	1.032	1.096	1.366	2.090	3.105	3.739	3.954	4.015
1	**0.5**	1.500	1.167	1.072	1.035	1.021	1.019	1.028	1.052	1.103	1.194
2	**1**	2.000	1.500	1.333	1.250	1.200	1.167	1.143	1.125	1.111	1.100
4	**2**	3.000	2.333	2.143	2.067	2.032	2.016	2.008	2.004	2.002	2.001

Figure 6. The amplitude ratio values from **Figures 123-126**. ω=0.01 s^{-1}, f=0.0016 s^{-1}.

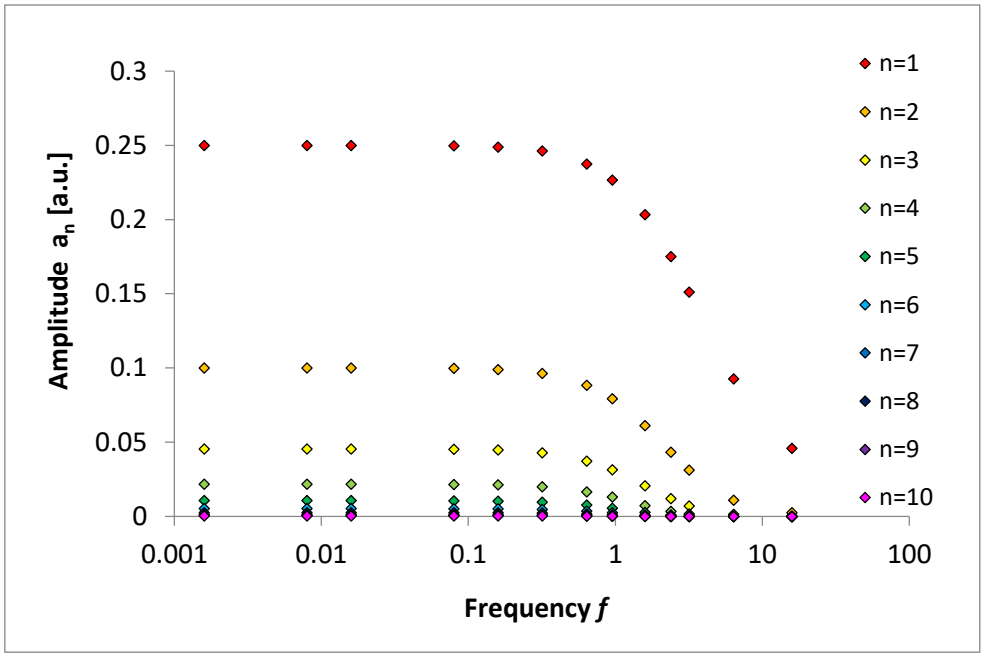


Figure 122. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f**=5** u/s^{-1}, C**=1** u, k_r**=10** s^{-1}, A_0**=1** u**,** B_0**=1** u, where u is an arbitrary unit of concentration. $\frac{k_r}{k_f} = 2\ u$, $\frac{k_r}{k_f C} = 2$.

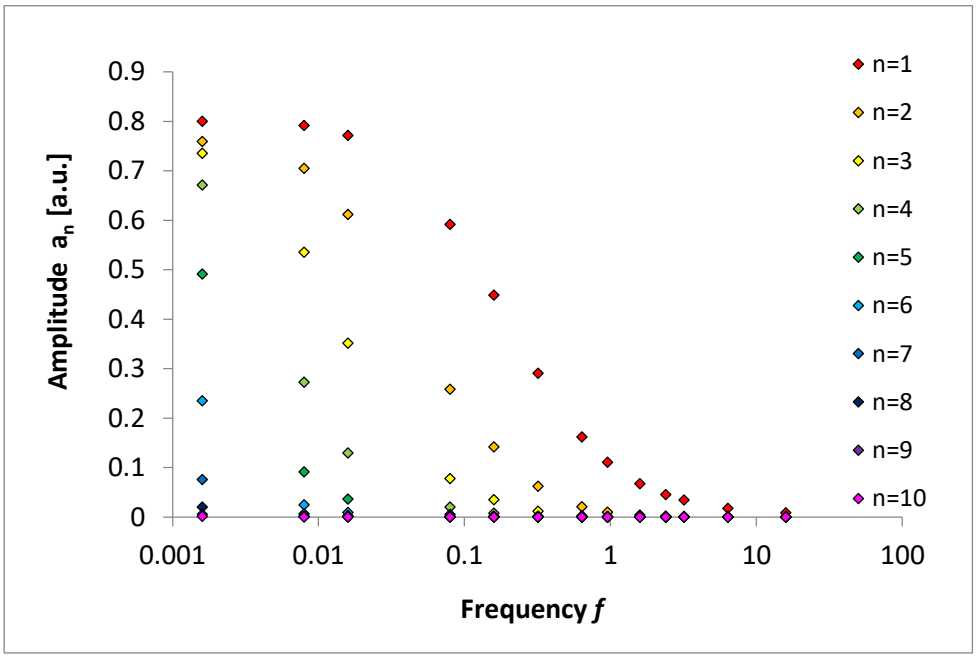


Figure 123. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f **=1** u/s^{-1}, C **=2** u, k_r **=0.5** s^{-1}, A_0**=1** u**,** B_0**=1** u, where u is an arbitrary unit of concentration. $\frac{k_r}{k_f} = 0.5\ u$, $\frac{k_r}{k_f C} = 0.25$.

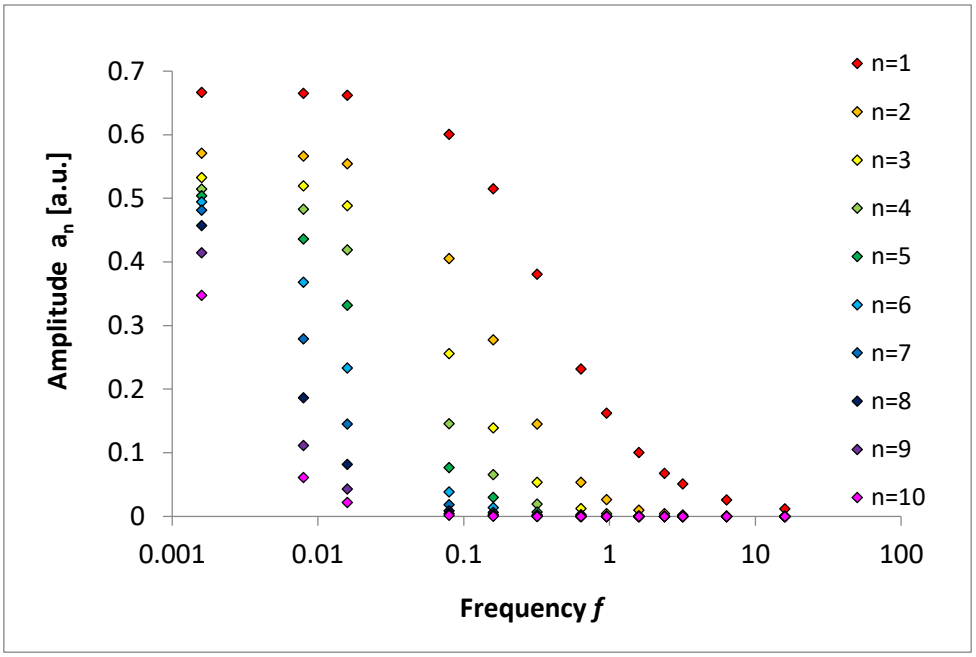


Figure 124. Strength of the signal propagated through a cascade as a function of its frequency, measured the amplitude of the *n*-numbered active form. k_f**=1** u/s^{-1}, C **=2** u, k_r **=1** s^{-1}, A_0 **=1** u, B_0 **=1** u, where *u* is an arbitrary unit of concentration. $\frac{k_r}{k_f} = 1\ u$, $\frac{k_r}{k_f C} = 0.5$.

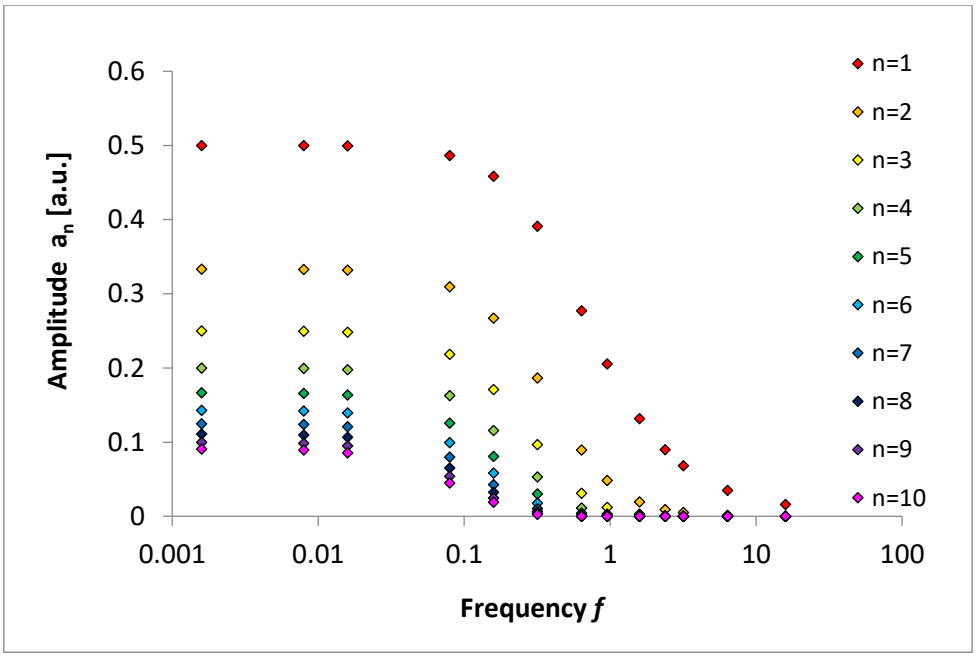


Figure 125. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f**=1** u/s^{-1}, C **=2** u, k_r **=2** s^{-1}, A_0 **=1** u, B_0 **=1** u, where *u* is an arbitrary unit of concentration. $\frac{k_r}{k_f} = 2\ u$, $\frac{k_r}{k_f C} = 1$.

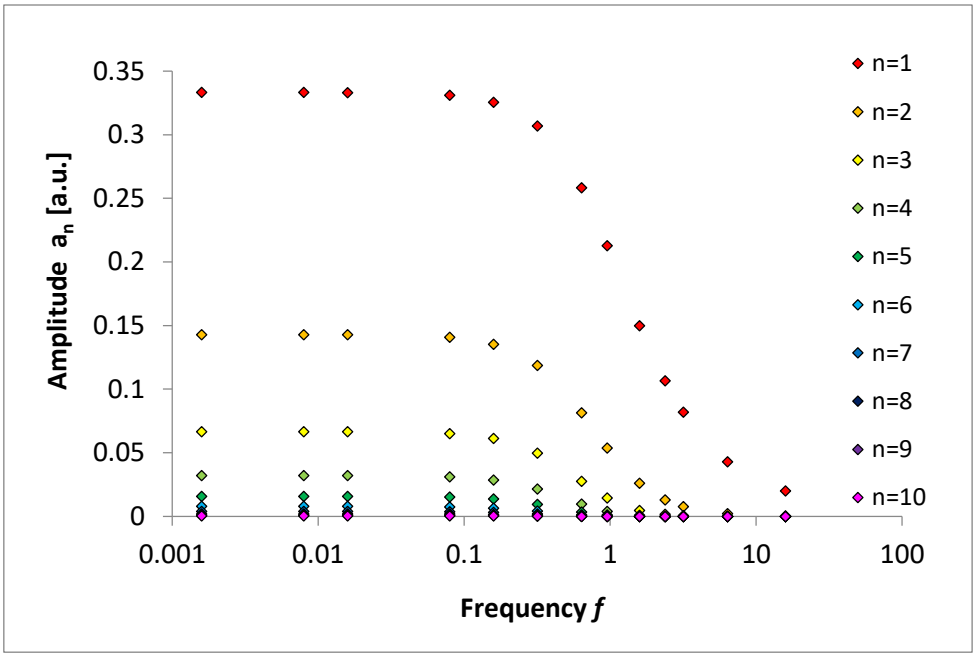


Figure 126. Strength of the signal propagated through a cascade as a function of its frequency, measured as the amplitude of the *n*-numbered active form. k_f**=1** u/s^{-1}, C**=2** u, k_r**=4** s^{-1}, A_0**=1** u, B_0**=1** u, where *u* is an arbitrary unit of concentration. $\frac{k_r}{k_f} = 4\ u$, $\frac{k_r}{k_f C} = 2$.

In summary, catalytic cascades exhibit two modes of operation. The first mode is in effect, when $\frac{k_{rev}}{k_{fw}C} > 1$ – in such a case, in the transmissive domain, the signal is strongly dampened after just a few layers, but the threshold frequency is apparent on the plots. The second mode is in effect, when $\frac{k_{rev}}{k_{fw}C} < 1$ – in such a case, in the transmissive domain, the signal maintains a reduced amplitude, allowing for a weak but long-range transmission, however the threshold frequency undergoes a significant shift at each consecutive layer, making the potential *a priori* predictions more challenging.

The value of $\frac{k_{rev}}{k_{fw}C}$ is dimensionless, because k_{fw} was measured as $\frac{concentration}{time}$, while k_{rev} as $\frac{1}{time}$. Thus, the $\frac{k_{rev}}{k_{fw}}$ has the dimension of concentration. This occurs because the "forward" reaction is catalyzed, while the "reverse" is spontaneous – e.g., for an individual layer *i* and a long time *t*:

$$\frac{d[a_i^{on}]}{dt} = k_{fw}[a_{i-1}^{on}][a_i^{off}] = \boldsymbol{k_{fw,eff,i}}[a_i^{off}], \qquad \boldsymbol{k_{fw,eff,i}} = k_{fw}[a_{i-1}^{on}]\ \left[\frac{1}{s}\right],$$

$$\frac{d[a_i^{off}]}{dt} = k_{rev}[a_i^{on}] = \boldsymbol{k_{rev,eff,i}}[a_i^{n}], \qquad \boldsymbol{k_{rev,eff,i}} = k_{rev}\ \left[\frac{1}{s}\right].$$

Based on the results from numeric simulations, I was finally able to construct equations for the effective rate constant of catalytic cascades. Since each cascade operates in one of two possible modes, there were two equations needed.

According to how, in some cases, the amplitudes decreased $\frac{k_{rev}}{k_{fw}C}$ or $\left(\frac{k_{rev}}{k_{fw}C}\right)^{-1}$-fold at each *n*-numbered layer, the equations that result in an overlap of the amplitude plots were expected to include either $\left(\frac{k_{rev}}{k_{fw}C}\right)^{n}$ or $\left(\frac{k_{rev}}{k_{fw}C}\right)^{-n}$. Now, since the effective rate constant k_{eff} is supposed to have a dimension of frequency ($1/_S$ or Hertz), and $\frac{k_{rev}}{k_{fw}C}$ itself is dimensionless, it had to be multiplied by some quantity with the desired dimension. The reverse rate constant k_{rev} was a natural first choice. Since the equations concern long homogeneous cascades, a simplification was allowed that:

$$[a_{i-1}^{on}] = C_{i-1} = C, \qquad \boldsymbol{k_{fw,eff,i}} = k_{fw}C\ \left[\frac{1}{s}\right],$$

which should be taken into consideration, when making predictions for short, heterogeneous cascades – in such cases:

$$\boldsymbol{k_{fw,eff,i}} = k_i^{fw}C_{i-1}, \qquad \boldsymbol{k_{fw,eff,i+1}} = k_{i+1}^{fw}C_i\ \left[\frac{1}{s}\right].$$

The first constructed candidates for an effective rate constant equation were:

$$\boldsymbol{k_{eff}} = \boldsymbol{k_{rev}} \cdot \left(\frac{\boldsymbol{k_{rev}}}{\boldsymbol{k_{fw}C}}\right)^{\boldsymbol{n}} \qquad \text{for} \qquad \frac{\boldsymbol{k_{rev}}}{\boldsymbol{k_{fw}C}} > \mathbf{1},$$

$$\boldsymbol{k_{eff}} = \boldsymbol{k_{rev}} \cdot \left(\frac{\boldsymbol{k_{rev}}}{\boldsymbol{k_{fw}C}}\right)^{\boldsymbol{-n}} \qquad \text{for} \qquad \frac{\boldsymbol{k_{rev}}}{\boldsymbol{k_{fw}C}} < \mathbf{1}.$$

When these equations were used, the plots of $A_n\left(\frac{f}{k_{eff,n}}\right)$ showed a remarkable, yet still imperfect overlap.

The constructed equations were further modified to improve the overlap and to have the apparent threshold domains located as close to $\frac{f}{k_{eff}} = 1$ as possible, ultimately resulting in the established equations of the cascade effective rate constants:

$$\boldsymbol{k_{eff} = k_{rev} \cdot \left(\frac{k_{rev}}{k_{fw}C}\right)^{2/n}} \qquad \text{for} \qquad \boldsymbol{\frac{k_{rev}}{k_{fw}C} > 1},$$

$$\boldsymbol{k_{eff} = k_{rev} \cdot \left(\frac{k_{rev}}{k_{fw}C}\right)^{n/2}} \qquad \text{for} \qquad \boldsymbol{\frac{k_{rev}}{k_{fw}C} < 1}.$$

By extension, a variant for heterogeneous cascades takes the form of the equations:

$$\boldsymbol{k_{eff} = \overline{k}_{rev} \cdot \prod_n \left(\frac{k_n^{rev}}{k_n^{fw} C_{n-1}}\right)^{\frac{2}{n^2}}} \qquad \text{for} \qquad \boldsymbol{\frac{k_{rev}}{k_{fw}C} > 1},$$

$$\boldsymbol{k_{eff} = \overline{k}_{rev} \cdot \prod_n \sqrt{\frac{k_n^{rev}}{k_n^{fw} C_{n-1}}}} \qquad \text{for} \qquad \boldsymbol{\frac{k_{rev}}{k_{fw}C} < 1}.$$

The established equations for the effective rate constant were tested against the biologically relevant model of an MAPK cascade, described in section **5.3**. The results are shown in **Figure 127**. Starting from *n*=2, the plots matched closely – which is most apparent when the amplitude plots are normalized. For *n*=1, the expressions applied poorly – which was to be expected as they were designed for entire networks and not for individual reactions. Besides, such reactions could be readily predicted from a steady-state approximation, as explained in sub-section **5.4.1**.

To further validate the established equations, the effective rate constants of all cascade models described in this section were found and the amplitudes were plotted against $\frac{f}{k_{eff}}$ – the results are available in sub-section **2.6.1**.

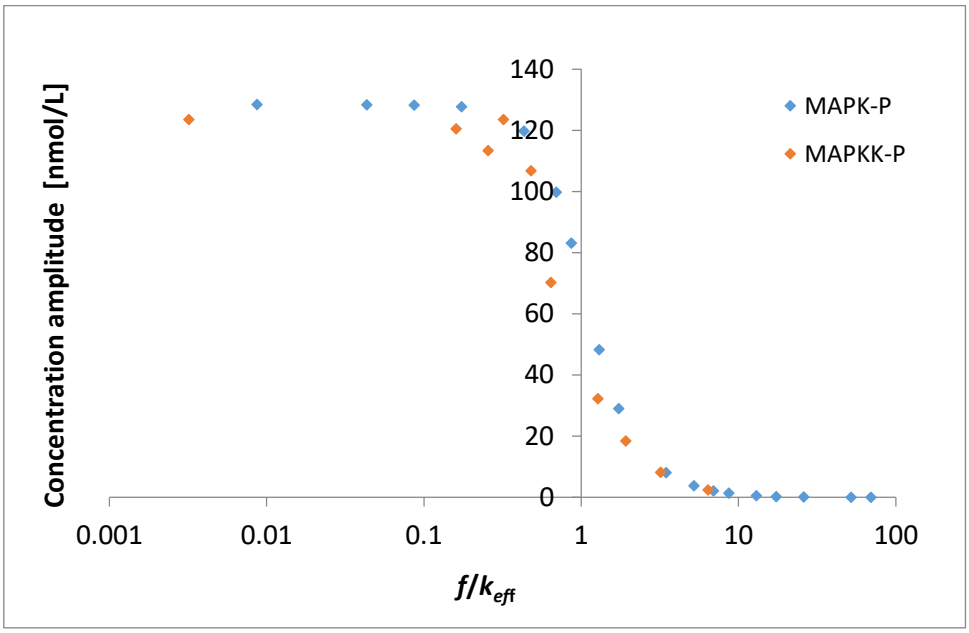


Figure 127. Strength of the propagated signal as a function of $\frac{f}{k_{eff}}$ in the cascades described in **Figure 75** and **Table 3**. Dampening of the signal begins near $\frac{f}{k_{eff}} = 1$.

5.5. Propagation of signals through realistic reaction networks – part 2

The methods established in section **5.4** allowed for a successful prediction of transmissive frequencies of a simple MAPK cascade network, described in section **5.1** – although this cascade featured enzymatic kinetics and slightly diverse parameters at each layer, it was very simplistic with its mono-phosphorylative activation and an arbitrary input signal. The complete validation required an analysis of the actual *in vivo* oscillations in a realistic network, both in terms of the architecture and parameters – such a network is presented in **Figure 128**, which illustrates a detailed model of the mammalian Ras/ERK kinase cascade. Contrary to its simplified predecessor, the cascade features the realistic reaction kinetics and plausible architecture, but most importantly it is modulated by the ubiquitous biological signal carried by the intracellular calcium oscillations – their observed physiological frequency ranging from 10^{-5} Hz (in neurons) up to 100 Hz (in skeletal muscle cells).[67]

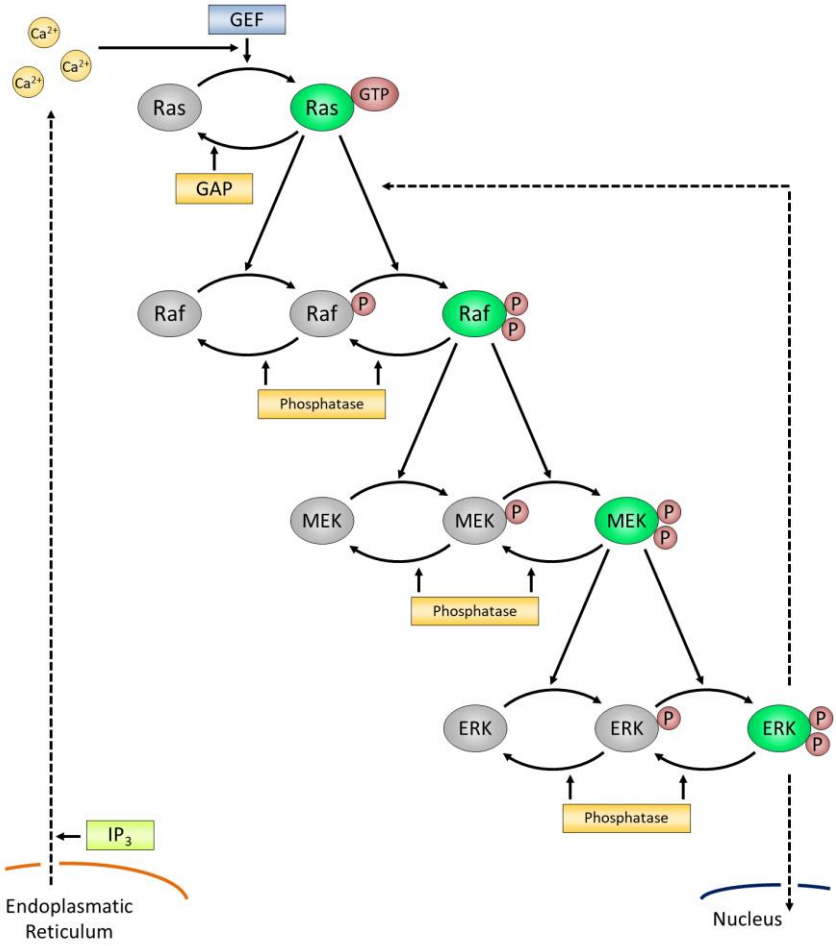


Figure 128. In this realistic model, the Ras/ERK cascade is activated by the guanine nucleotide exchange factor (GEF, the blue square), whose activity is modulated by the waves of Ca^{2+} concentration (released from the endoplasmic reticulum in a pulsating manner), described as $[\mathrm{Ca}^{2+}] = 0.45\sin(\omega t) + 0.55\ [\mu\mathrm{M}]$. The concentration of the GEF itself is considered constant. The grey ovals denote the inactive kinases, whereas the green ovals – the kinases activated via dual phosphorylation. The total concentration of Ras is constant at 100 nM, while the total concentrations of Raf, MEK and ERK are each equal to 300 nM.[59] Each active kinase catalyzes the activation of another kinase in the subsequent layer. In the case of Ras, activation requires only one phosphorylation event, while in the case of Raf, MEK and ERK, it required two subsequent phosphorylations. The activity of the enzymes that deactivate kinases (the orange squares) is constant. The output signal is identified as the momentary concentration of ERK-PP, which sends the signal forth to the nucleus and modulates the activity of Ras in a feedback loop, either amplifying or decreasing the activation rate of Raf according to the *F* parameter. The feedback is positive for $F > 1$, negative for $F < 1$, and null when $F = 1$.

The network itself was assembled from two existing models – the main segment, resembling a typical MAPK assembly introduced in **Figure 10**, which came from the works of Kholodenko,[59] and the uppermost layer, responsible for decoding of the calcium signal, which came from the works of Yi.[73] The input signal was set to remain within typical cellular Ca^{2+} levels[28,68] of 100 nM to 1µM. According to the convention established in section **5.1**, the network is described by the following equations:

$$RAS \underset{(2)}{\overset{Ca^{2+},\ f(ERK-PP)\ (1)}{\rightleftarrows}} RAS-GTP,$$

$$RAF \underset{(6)}{\overset{RAS-GTP\ (3)}{\rightleftarrows}} RAF-P \underset{(5)}{\overset{RAS-GTP\ (4)}{\rightleftarrows}} RAF-PP,$$

$$MEK \underset{(10)}{\overset{RAF-PP\ (7)}{\rightleftarrows}} MEK-P \underset{(9)}{\overset{RAF-PP\ (8)}{\rightleftarrows}} MEK-PP,$$

$$ERK \underset{(14)}{\overset{MEK-PP\ (11)}{\rightleftarrows}} ERK-P \underset{(13)}{\overset{MEK-PP\ (12)}{\rightleftarrows}} ERK-PP.$$

The respective rate equations are presented in **Table 7** – there, the equations (3) and (4) are the most discernible, as each comprises a product of the Michaelis-Menten equation and an expression of feedback from ERK-PP.

All the simulations were performed with one of three values of F: 1 (no feedback), 0.34 (a strong feedback) or 0.01 (a very strong feedback). While it was possible to model a positive feedback, it was highly implausible to exist *in vivo* – in the original works of Kholodenko it was used to study the bifurcation behavior of the ERK cascade, connected to the emergence of oscillations caused by the feedback loop.[59]

No.	Reaction	Rate expression	Kinetic constants
1	GEF Ras → Ras-GTP	$v_{fw}^{Ras} = \left(V_{basal}^{Ras} + \frac{k_{fw}^{Ras}[Ca^{2+}]^2}{K_M^{Ras\,2} + [Ca^{2+}]^2}\right)[Ras]$	$V_{basal}^{Ras} = 0.001\ \mathrm{s}^{-1}$ $k_{fw}^{Ras} = 2\ \mathrm{s}^{-1}$ $K_M^{Ras} = 1\ \mu\mathrm{M}$
2	GAP Ras-GTP → Ras	$v_{rev}^{Ras} = V_{rev}^{Ras}[Ras\_GTP]$	$V_{rev}^{Ras} = 0.6\ \mathrm{s}^{-1}$
3	Ras GTPase 1 Raf → Raf-P	$v_{fw1}^{Raf} = \frac{k_{fw1}^{Raf}[Ras\_GTP]\,[Raf]/K_{M1}^{Raf}}{1 + [Raf]/K_{M1}^{Raf} + [Raf\_P]/K_{M2}^{Raf}} \cdot \frac{1 + F[ERK\_PP\,]/K_f}{1 + [ERK\_PP\,]/K_f}$	$k_{fw1}^{Raf} = 1\ \mathrm{s}^{-1}$ $K_{M1}^{Raf} = 100\ \mathrm{nM}$ $K_{M2}^{Raf} = 200\ \mathrm{nM}$ $K_f = 25\ \mathrm{nM}$
4	Ras GTPase 2 Raf-P → Raf-PP	$v_{fw2}^{Raf} = \frac{k_{fw2}^{Raf}[Ras\_GTP]\,[Raf]/K_{M2}^{Raf}}{1 + [Raf]/K_{M1}^{Raf} + [Raf\_P]/K_{M2}^{Raf}} \cdot \frac{1 + F[ERK\_PP\,]/K_f}{1 + [ERK\_PP\,]/K_f}$	$k_{fw2}^{Raf} = 0.25$
5	Raf phosphatase 1 Raf-PP → Raf-P	$v_{rev1}^{Raf} = \frac{V_{rev1}^{Raf}\,[Raf\_PP]/K_{Mp1}^{Raf}}{1 + [Raf\_PP]/K_{Mp1}^{Raf} + [Raf\_P]/K_{Mp2}^{Raf}}$	$V_{rev1}^{Raf} = 2.5\ \mathrm{s}^{-1}$ $K_{Mp1}^{Raf} = 50\ \mathrm{nM}$ $K_{Mp2}^{Raf} = 100\ \mathrm{nM}$
6	Raf phosphatase 2 Raf-P → Raf	$v_{rev2}^{Raf} = \frac{V_{rev2}^{Raf}\,[Raf\_P]/K_{Mp2}^{Raf}}{1 + [Raf\_PP]/K_{Mp1}^{Raf} + [Raf\_P]/K_{Mp2}^{Raf}}$	$V_{rev2}^{Raf} = 3.75\ \mathrm{s}^{-1}$
7	Raf kinase 1 MEK → MEK-P	$v_{fw1}^{MEK} = \frac{k_{fw1}^{MEK}[Raf\_PP]\,[MEK]/K_{M1}^{MEK}}{1 + [MEK]/K_{M1}^{MEK} + [MEK\_P]/K_{M2}^{MEK}}$	$k_{fw1}^{MEK} = 2.5\ \mathrm{s}^{-1}$ $K_{M1}^{MEK} = 250\ \mathrm{nM}$ $K_{M2}^{MEK} = 250\ \mathrm{nM}$
8	Raf kinase 2 MEK-P → MEK-PP	$v_{fw2}^{MEK} = \frac{k_{fw2}^{MEK}[Raf\_PP]\,[MEK\_P]/K_{M2}^{MEK}}{1 + [MEK]/K_{M1}^{MEK} + [MEK\_P]/K_{M2}^{MEK}}$	$k_{fw2}^{MEK} = 0.5\ \mathrm{s}^{-1}$
9	MEK phosphatase 1 MEK-PP → MEK-P	$v_{rev1}^{MEK} = \frac{V_{rev1}^{MEK}\,[MEK\_PP]/K_{Mp1}^{MEK}}{1 + [MEK\_PP]/K_{Mp1}^{MEK} + [MEK\_P]/K_{Mp2}^{MEK}}$	$V_{rev1}^{MEK} = 3\ \mathrm{s}^{-1}$ $K_{Mp1}^{MEK} = 250\ \mathrm{nM}$ $K_{Mp2}^{MEK} = 80\ \mathrm{nM}$
10	MEK phosphatase 2 MEK-P → MEK	$v_{rev2}^{MEK} = \frac{V_{rev2}^{MEK}\,[MEK\_P]/K_{Mp2}^{MEK}}{1 + [MEK\_PP]/K_{Mp1}^{MEK} + [MEK\_P]/K_{Mp2}^{MEK}}$	$V_{rev2}^{MEK} = 3.75\ \mathrm{s}^{-1}$
11	MEK kinase 1 ERK → ERK-P	$v_{fw1}^{ERK} = \frac{k_{fw1}^{ERK}[MEK\_PP]\,[ERK]/K_{M1}^{ERK}}{1 + [ERK]/K_{M1}^{ERK} + [ERK\_P]/K_{M2}^{ERK}}$	$k_{fw1}^{MEK} = 0.125\ \mathrm{s}^{-1}$ $K_{M1}^{MEK} = 250\ \mathrm{nM}$ $K_{M2}^{MEK} = 250\ \mathrm{nM}$
12	MEK kinase 2 ERK-P → ERK-PP	$v_{fw2}^{ERK} = \frac{k_{fw2}^{ERK}[MEK\_PP]\,[ERK\_P]/K_{M2}^{ERK}}{1 + [ERK]/K_{M1}^{ERK} + [ERK\_P]/K_{M2}^{ERK}}$	$k_{fw2}^{MEK} = 0.125\ \mathrm{s}^{-1}$
13	ERK phosphatase 1 ERK-PP → ERK-P	$v_{rev1}^{ERK} = \frac{V_{rev1}^{ERK}\,[ERK\_PP]/K_{Mp1}^{ERK}}{1 + [ERK\_PP]/K_{Mp1}^{ERK} + [ERK\_P]/K_{Mp2}^{ERK} + [ERK]/K_{Mp3}^{ERK}}$	$V_{rev1}^{ERK} = 3.75\ \mathrm{s}^{-1}$ $K_{Mp1}^{ERK} = 120\ \mathrm{nM}$ $K_{Mp2}^{ERK} = 20\ \mathrm{nM}$ $K_{Mp3}^{ERK} = 300\ \mathrm{nM}$
14	ERK phosphatase 2 ERK-P → ERK	$v_{rev2}^{ERK} = \frac{V_{rev2}^{ERK}\,[ERK\_P]/K_{Mp2}^{ERK}}{1 + [ERK\_PP]/K_{Mp1}^{ERK} + [ERK\_P]/K_{Mp2}^{ERK} + [ERK]/K_{Mp3}^{ERK}}$	$V_{rev2}^{ERK} = 5\ \mathrm{s}^{-1}$

Table 7. The equations governing the Ras/ERK cascade from **Figure 128**. The input signal is $[\mathrm{Ca}^{2+}] = 0.45\sin(\omega t) + 0.55\ [\mu\mathrm{M}]$, the total concentration of Ras is 100 nM and the total concentrations of Raf, MEK and ERK are 300 nM each.

In order to estimate the effective reaction rate constant k_{eff} of the entire Ras/ERK cascade, it was necessary to determine its architecture first – a task which was not so trivial since, from the point of this study, the network was not a "pure" cascade, but a combination of various motifs – including the feedback loop spanning across the network. The cascade assembly was the most apparent, but three of its layers featured multiple phosphorylation cycles, in some way resembling the reversible linear sequences from sections **5.1**-**5.3** – and then, there was the feedback loop connecting the opposite ends of the cascade.

In order to solve this network, I decided to "disassemble" it by identifying the effective rate constants of the underlying components first – once such "local" k_{eff} were found, their respective elements were effectively removed from the network, reducing it to a solvable form. The process of this iterative "removal" is specifically illustrated in **Figure 129a**. First, the feedback loop was excluded, which proved to be a straightforward operation – according to sub-section **5.4.1**, the k_{eff} of the affected reactions (3) and (4) were estimated from the steady-state approximation, as described in sub-section **5.4.1**:

$$\boldsymbol{k_{eff,(3)}} \approx \frac{k_{fw1}^{Raf}[Ras_{GTP}]/K_{M1}^{Raf}}{1+[Raf]/K_{M1}^{Raf}+[Raf_P]/K_{M2}^{Raf}} \cdot \frac{1+F[ERK_{PP}]/K_f}{1+[ERK_{PP}]/K_f},$$

$$\boldsymbol{k_{eff,(4)}} \approx \frac{k_{fw2}^{Raf}[Ras_{GTP}]/K_{M2}^{Raf}}{1+[Raf]/K_{M1}^{Raf}+[Raf_P]/K_{M2}^{Raf}} \cdot \frac{1+F[ERK_{PP}]/K_f}{1+[ERK_{PP}]/K_f}.$$

As a result, these two constants included the feedback in its entirety, so it could be omitted in the following steps. The layers that comprised two phosphorylation cycles were disassembled into two linear reaction sequences each – as such they could be solved for k_{eff} according to the methods described in section **5.3**. This arguably convoluted process of layer disassembly is presented step-by-step in **Figure 129b**. After all these operations, the network began to resemble the model cascade from sub-section **5.4.2** and was solved accordingly, giving one effective rate constant that comprises the entire Ras/ERK pathway.

The subsequent parts of this section describe the mathematics behind the cascade "disassembly". Depending on the value of the $\frac{k_{rev}}{k_{fw}C}$ ratio, one of two k_{eff} equations may apply to the cascade. Since the Ras/ERK cascade features a significant diversity of its layers, each layer might qualify for different equation:

$$\boldsymbol{k_{eff} = \bar{k}_{rev} \cdot \prod_n \sqrt{\frac{k_n^{rev}}{k_n^{fw} C_{n-1}}}} \qquad \text{for} \qquad \boldsymbol{\frac{k_{rev}}{k_{fw}C} < 1},$$

$$\boldsymbol{k_{eff} = \bar{k}_{rev} \cdot \prod_n \left(\frac{k_n^{rev}}{k_n^{fw} C_{n-1}}\right)^{\frac{2}{n^2}}} \qquad \text{for} \qquad \boldsymbol{\frac{k_{rev}}{k_{fw}C} > 1}.$$

For inhibition factors F=1 and F=0.34, $\boldsymbol{\frac{k_{rev}}{k_{fw}C} < 1}$ in the Raf, MEK and ERK layers, but $\boldsymbol{\frac{k_{rev}}{k_{fw}C} > 1}$ in the Ras layer. However, for inhibition factor F=0.01, $\boldsymbol{\frac{k_{rev}}{k_{fw}C} < 1}$ in the MEK and ERK layers, while $\boldsymbol{\frac{k_{rev}}{k_{fw}C} > 1}$ in the Ras and Raf layers. With $\bar{k}_{rev}$ being the geometric average, the assembled equations are as follows:

$$\boldsymbol{k_{eff} = \sqrt[4]{k_{rev}^{Ras} k_{rev}^{Raf} k_{rev}^{MEK} k_{rev}^{ERK}} \cdot \left(\frac{k_{rev}^{Ras}}{k_{fw}^{Ras} C_{Ca^{2+}}}\right)^2 \cdot \sqrt{\frac{k_{rev}^{Raf}}{k_{fw}^{Raf} C_{Ras}}} \cdot \sqrt{\frac{k_{rev}^{MEK}}{k_{fw}^{MEK} C_{Raf}}} \cdot \sqrt{\frac{k_{rev}^{ERK}}{k_{fw}^{ERK} C_{MEK}}}} \qquad \left[\frac{1}{s}\right]$$

$$\text{for} \qquad \boldsymbol{F = 1} \text{ or } \boldsymbol{0.34},$$

$$\boldsymbol{k_{eff} = \sqrt[4]{k_{rev}^{Ras} k_{rev}^{Raf} k_{rev}^{MEK} k_{rev}^{ERK}} \cdot \sqrt{\frac{k_{rev}^{Ras}}{k_{fw}^{Ras} C_{Ca^{2+}}} \cdot \frac{k_{rev}^{Raf}}{k_{fw}^{Raf} C_{Ras}}} \cdot \sqrt{\frac{k_{rev}^{MEK}}{k_{fw}^{MEK} C_{Raf}}} \cdot \sqrt{\frac{k_{rev}^{ERK}}{k_{fw}^{ERK} C_{MEK}}}} \qquad \left[\frac{1}{s}\right]$$

$$\text{for} \qquad \boldsymbol{F = 0.01},$$

where all *k* are effective rate constants of the respective network components. The total concentrations *C* are 300 nM for Raf, MEK and ERK and 100 nM for Ras. The value of $\boldsymbol{C_{Ca^{2+}}}$ is 0.55 µM.

According to the general equation $1/k_{eff,n}^2 = 0.5 \sum_1^n 1/k_{eff,i}^2$ the expressions for the Raf, MEK and ERK layers (one "forward" and one "backward" for each layer) are as follows:

$$\boldsymbol{k_{fw}^{Raf}} = \left(\mathbf{0.5} \cdot \left(\frac{\mathbf{1}}{\boldsymbol{k_{Ras-GTPase\_1}^{eff}}^{\mathbf{2}}} + \frac{\mathbf{1}}{\boldsymbol{k_{Ras-GTPase\_2}^{eff}}^{\mathbf{2}}} \right) \right)^{-\frac{\mathbf{1}}{\mathbf{2}}},$$

$$\boldsymbol{k_{rev}^{Raf}} = \left(\mathbf{0.5} \cdot \left(\frac{\mathbf{1}}{\boldsymbol{k_{Raf\_phosphatase\_1}^{eff}}^{\mathbf{2}}} + \frac{\mathbf{1}}{\boldsymbol{k_{Raf\_phosphatase\_2}^{eff}}^{\mathbf{2}}} \right) \right)^{-\frac{\mathbf{1}}{\mathbf{2}}},$$

$$\boldsymbol{k_{fw}^{MEK}} = \left(\mathbf{0.5} \cdot \left(\frac{\mathbf{1}}{\boldsymbol{k_{Raf\_kinase\_1}^{eff}}^{\mathbf{2}}} + \frac{\mathbf{1}}{\boldsymbol{k_{Raf\_kinase\_2}^{eff}}^{\mathbf{2}}} \right) \right)^{-\frac{\mathbf{1}}{\mathbf{2}}},$$

$$\boldsymbol{k_{rev}^{MEK}} = \left(\mathbf{0.5} \cdot \left(\frac{\mathbf{1}}{\boldsymbol{k_{MEK\_phosphatase\_1}^{eff}}^{\mathbf{2}}} + \frac{\mathbf{1}}{\boldsymbol{k_{MEK\_phosphatase\_2}^{eff}}^{\mathbf{2}}} \right) \right)^{-\frac{\mathbf{1}}{\mathbf{2}}},$$

$$\boldsymbol{k_{fw}^{ERK}} = \left(\mathbf{0.5} \cdot \left(\frac{\mathbf{1}}{\boldsymbol{k_{MEK\_kinase\_1}^{eff}}^{\mathbf{2}}} + \frac{\mathbf{1}}{\boldsymbol{k_{MEK\_kinase\_2}^{eff}}^{\mathbf{2}}} \right) \right)^{-\frac{\mathbf{1}}{\mathbf{2}}},$$

$$\boldsymbol{k_{rev}^{ERK}} = \left(\mathbf{0.5} \cdot \left(\frac{1}{k_{ERK\_phosphatase\_1}^{eff}{}^{2}} + \frac{1}{k_{ERK\_phosphatase\_2}^{eff}{}^{2}} \right) \right)^{-\frac{\mathbf{1}}{\mathbf{2}}}.$$

As described in sub-section **5.4.1**, the effective kinetic rate constants of individual reactions involved in layers were obtained from a steady state approximation.

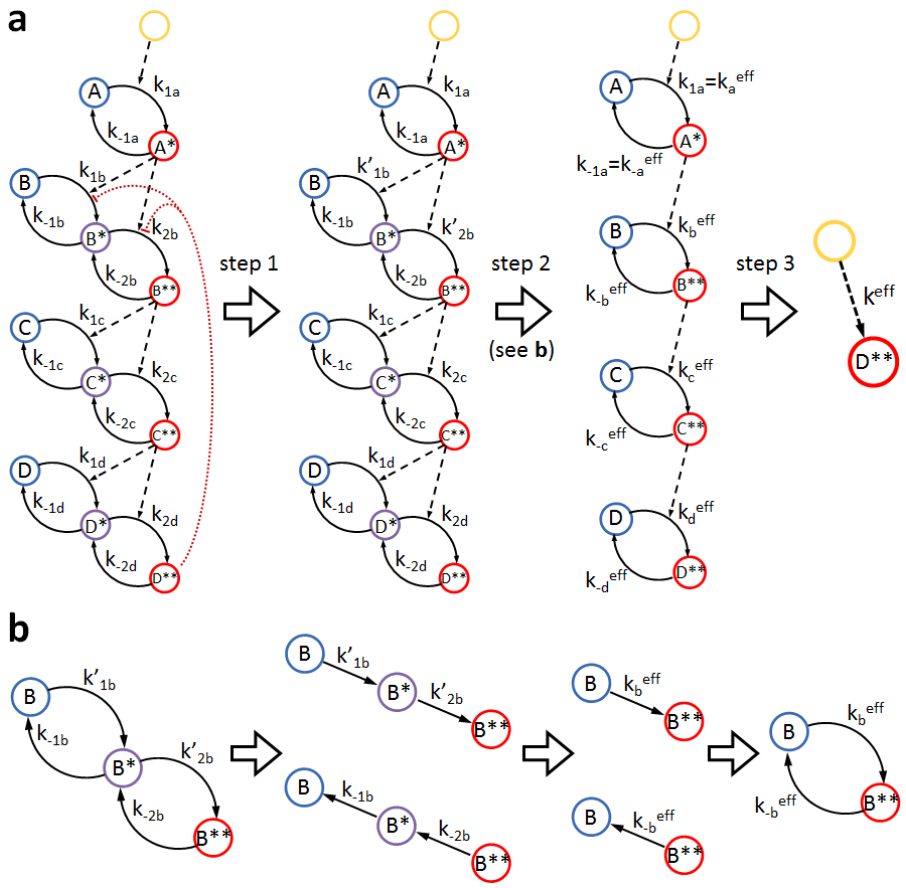


Figure 129. Step-wise deconstruction of the Ras/ERK cascade into simpler elements. The network from **Figure 128** is redrawn in (a). The dotted arrows indicate catalysis, the solid arrows indicate transformations, the asterisk indicates phosphorylation, the orange circle is calcium and D** is ERK-PP. This network is simplified to find the final value of k_{eff} between the first and the last nodes. In the first step, the feedback loop (marked with the burgundy dotted lines) is assumed as a factor already included in constants k_{1b} and k_{2b} and can be omitted in the further analysis. To annotate this, the rate constants are written with apostrophes after the first step, but their values remain the same. In the second step, each of the three layers (B, C and D) are simplified to single loops, each with two effective rate constants, forward and backward. Once the complex layers are simplified, the network has the same structure as the cascade presented in **Figure 81**. b) In order to simplify the two-loop layer, it is divided into two linear subsystems: one comprises the forward reactions described by k'_{1b} and k'_{2b}, and the other, reverse reactions described by k_{-1b} and k_{-2b}. The effective rate constant, k_{eff}, for each subsystem is found as for linear chains, and these values are used in step (3) of panel a) to ultimately approximate k_{eff}, of the entire cascade. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.[20]

6. Results and discussion

The aim of this work has been achieved. Cyclorg allowed for the discovery of potential alternative biochemistries and useful chemical systems by identifying numerous cyclic patterns in the Network of Organic Chemistry. Kinetix allowed for the discovery and analysis of an emergent property of (bio)chemical cycle cascades. Both pieces of the software have been improved and released for public use[4,8] – when used in tandem, they could potentially allow for qualitative design of chemical cycles and then quantitative modelling of their reaction kinetics.

The assortment of the discovered cycles includes chemically plausible candidates featuring an auto-amplification (an emergent property of interest to systems chemists), biomimesis (relevant to synthetic biology and origins of life) and the ability to recycle the reagents used (desired in green chemistry, as well as useful for industrial purposes). The best examples have been described both in this thesis and in *Angewandte Chemie*.[19]

The databases used for the cycle identification have proved to be the most significant drawback of the presented methodology. The lack of consistent data on the reaction conditions and yields made it impossible to automatically sort the identified cycles by chemical plausibility. In order to minimize the effect of this inherent problem, I have designed a set of reactant-based, hard-coded filters and optional sorting tools. I have also devised an optimal setup of search parameters, which (if selected by the user) further mitigate the inefficiency caused by the databases. However, a new method is required if chemical cycles are to be designed/discovered reliably.

The analysis of the cycle candidates (both good and bad) identified from the NOC allows for conceptualization of potential new methods of the cycle discovery, including the *de novo* tools. Many examples (the negative ones) could also allow for recognizing beforehand the problems that might arise during the development of such methods.

The kinetic studies of chemical systems in a stationary state have shown that every chemical reaction is characterized by a specific cut-off frequency k_{eff} (named effective rate constant) – all the concentration oscillations of frequency significantly lower than k_{eff} are fully transmitted by that reaction, while those of significantly higher frequency are completely dampened. Therefore, every reaction can be compared to an electric low-pass filter which transforms alternating current of sufficiently high frequency into direct current. By extension, entire chemical networks can be characterized by one cut-off frequency k_{eff} that comprises the entirety of their reaction kinetics – thus, is it possible to predict *a priori* (i.e., without numerical calculations) if such a system is capable of transmitting a signal of a given frequency. Not only this knowledge is relevant to systems chemistry and synthetic biology, but also to chemical engineering as it allows for predicting the maximum interval with which a substrate can be safely loaded into a chemical reactor.

In case of a single first order reaction, k_{eff} equals its rate constant. For higher-order reactions, k_{eff} must be determined according to the procedure described in sub-section **5.4.1**. In case of reaction networks, the general solution for k_{eff} depends on their architecture.

The simplest network comprises a linear sequence of reactions, with an initial substrate being fed in an oscillatory manner and the final product being removed continuously. In such a system, the amplitude change at any step is expressed as:

$$\frac{A_{n+1}}{A_n} = \frac{k_n}{\sqrt{{k_{n+1}}^2 + (2\pi f)^2}},$$

where *A* is an amplitude, *n* is a step number, *k* is an effective rate constant of an individual reaction and *f* is a signal frequency. The cut-off frequency (effective constant) is measured in $1/_S$ and expressed as:

$$\frac{1}{{k_{eff,n}}^2} = 0.5 \sum_{i=1}^{n} \frac{1}{{k_{eff,i}}^2},$$

where $k_{eff,n}$ is a network effective constant, $k_{eff,i}$ is effective constant of an individual *i*-numbered reaction. See **Figure 72** or **130a** for validation.

Catalytic cascades have two modes of operation – one where low-frequency oscillations are fully transmitted short-range, but strongly dampened after few initial steps and another where low-frequency signals can be transmitted long-range, but at the cost of reduced strength (details in sub-section **5.4.2.**). Kinetix allowed for numerical solving of the cut-off frequency (always measured in $^1/_S$) of the cascades, which could not be solved analytically:

$$\boldsymbol{k_{eff} = \bar{k}_{rev} \cdot \prod_n \sqrt{\frac{k_n^{rev}}{k_n^{fw} C_{n-1}}}} \qquad \text{for} \qquad \boldsymbol{\frac{k_{rev}}{k_{fw} C} < 1}, \text{ (long range)}$$

$$\boldsymbol{k_{eff} = \bar{k}_{rev} \cdot \prod_n \left(\frac{k_n^{rev}}{k_n^{fw} C_{n-1}}\right)^{\frac{2}{n^2}}} \qquad \text{for} \qquad \boldsymbol{\frac{k_{rev}}{k_{fw} C} > 1}, \text{ (short range)}$$

where $\boldsymbol{k_n^{fw}}$ is an effective rate constant of activation at the *n*-th step, $\boldsymbol{k_n^{rev}}$ is an effective rate constant of deactivation at the *n*-th step, $\boldsymbol{\bar{k}_{rev}}$ is a geometric average of all deactivation constants and $\boldsymbol{C_n}$ is a total concentration of all forms in the *n*-th cycle. The equations are more accurate for less diverse values at each layer. See **Figure 130b,c** for validation.

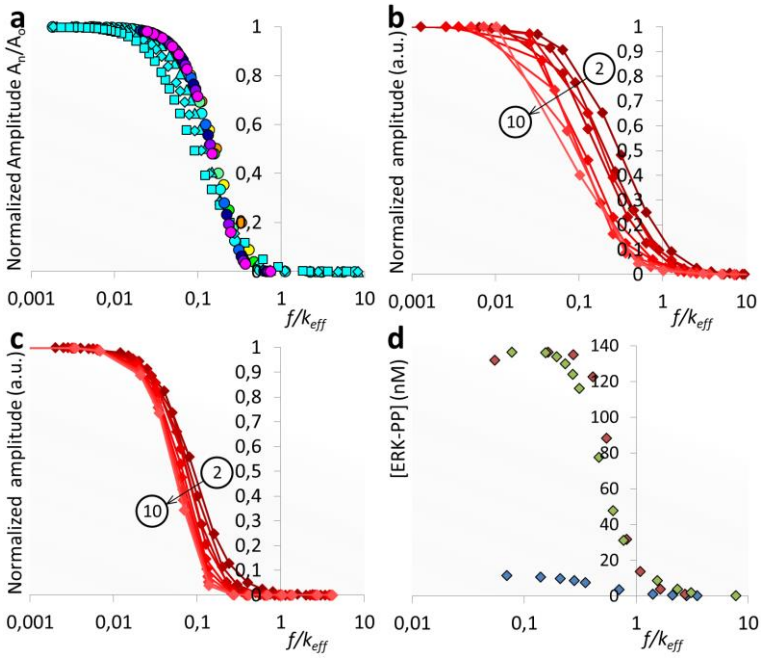


Figure 130. Signal propagation along the chemical networks: (a) at linear chains with varying number of the 1st order or reversible reactions (from 1 to 10), where the forward rate constant is $1s^{-1}$ at each step and reverse varies from 0 to $0.75s^{-1}$, (b) at each of the 10 layers of the cascade featuring the 1st order kinetics, where k_{fw}=1, k_{rev}=2, C=1 at every layer, (c) at each of the 10 layers of cascade featuring the 1st order kinetics, where k_{fw}=2, k_{rev}=1, C=1 at every layer and (d) in the realistic Ras/ERK cascade (blue = the 1st layer, red = the 2nd layer, green = 3rd layer). Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.[20]

Two biologically-relevant models have been investigated using Kinetix (described in **5.3.** and **5.5.**) Their cut-off frequencies have been successfully determined using the previously presented equations. The simple cascade shown in **5.3.** has k_{eff} equal to 0.18 Hz for the MAPK (third) layer and 0.50 Hz for the MAPKK (second) layer, see **Figure 127** (presented for the second time on this page for convenience purposes).

The analysis of the Ras/ERK cascade from **5.5.** has shown that k_{eff} can be determined for a complex network comprising diverse architectures by disassembling it into subsystems and solving each separately. The cut-off frequencies are 0.21 s^{-1} for the Raf layer, 0.014 s^{-1} for the MEK layer and 0.00045 s^{-1} for the ERK layer – see **Figure 130d**.

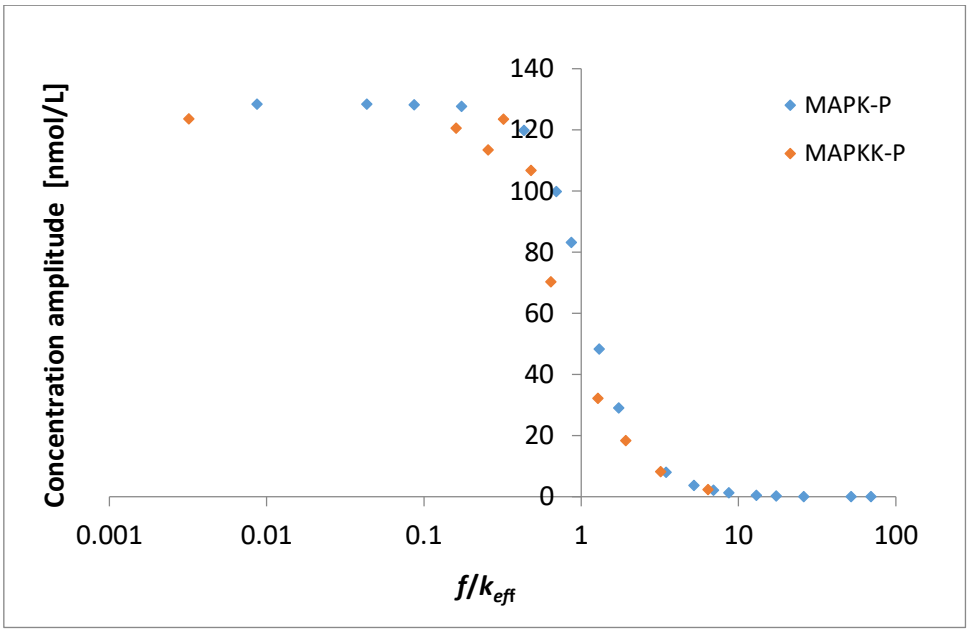


Figure 127. Strength of the propagated signal as a function of $\frac{f}{k_{eff}}$ in the cascade described in **Figure 75** and **Table 3**. Dampening of the signal begins near $\frac{f}{k_{eff}} = 1$. MAPKKK-P is not shown, because as the first layer it is better described as a single reversible reaction rather than a cascade.

Lastly, Kinetix software has allowed for the discovery and analysis of the mechanism by which (bio)chemical systems transform time-varying signals (e.g., travelling waves of ion concentration) into a continuous response, as presented in **Figure 131**. Here, as calcium frequency approaches the network effective constant, its amplitude decreases, but at the same time the average concentration increases – resulting in continuous activation of the factors targeted by the cascade. This is an emergent property of signaling cascades, which is modulated by the total concentration of the kinase at each layer, and might be another reason for their involvement in cellular signaling.

The frequencies of the model Ras/ERK activation have been compared to the *in vitro* experimental results. With the inhibitory feedback present, the cascade dampens the oscillations with frequencies above the ~0.001-0.004 Hz threshold, at the same time increasing the strength of the continuous signal. This is remarkably close to the transmission-dampening transition recorded experimentally[74] at ~ 0.0083-0.016 Hz for the Ras/ERK pathway.

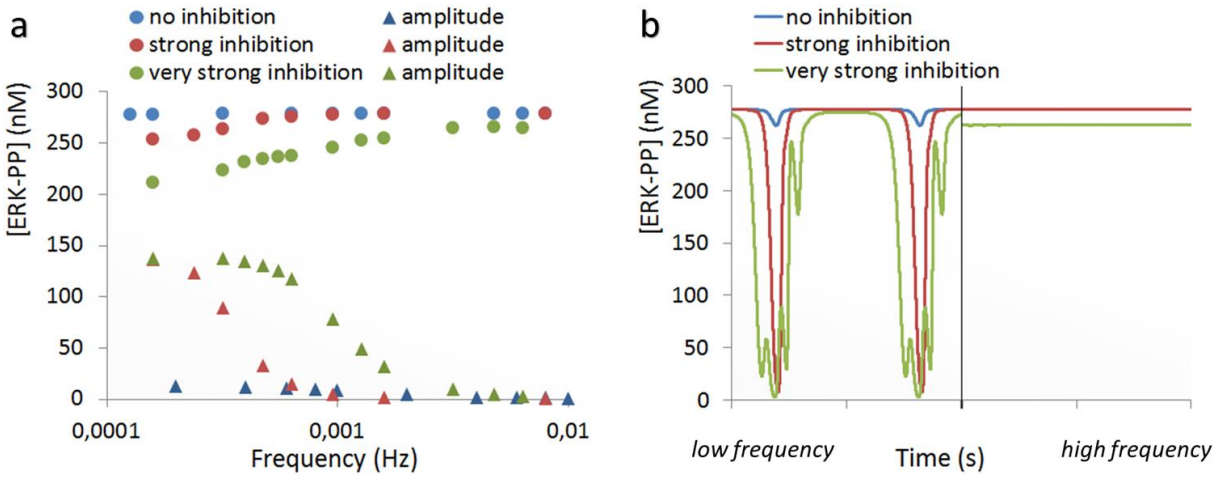


Figure 131. Numerical analysis of the Ras/ERK model. a) Full frequency dependence ERK-PP amplitude (the solid triangles) and average concentration (the filled circles). b) Typical concentration profiles of ERK-PP for the low- (left part, $f=1.59\cdot10^{-4}$ Hz) and high-frequency (right part, $f=1.59\cdot10^{-2}$ Hz) calcium inputs. The feedback is described with parameter F = 1 (blue), F = 0.34 (red), F = 0.01 (green).

7. References

[1] B. A. Grzybowski, K. J. M. Bishop, B. Kowalczyk, C. E. Wilmer, The 'wired' universe of organic chemistry. *Nat. Chem.*, **2009**, 1, 31-36

[2] R. Sedgewick, Algorithms in C, Part 5: Graph Algorithms, 3rd ed., Addison Wesley Professional, Reading, **2001**

[3] A. A. Hagberg, D. A. Schult, P. J. Swart, *Proceedings of the 7th Python in Science Conference (SciPy2008)*, 11-15, **2008**

[4] http://cyclorg.grzybowskigroup.pl

[5] https://www.meteor.com/

[6] E. Süli, D. F. Mayers, *An Introduction to Numerical Analysis*, Cambridge University Press, Cambridge, **2003**

[7] J. Crank, P. Nicolson, A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Math. Proc. Cambridge Phisol. Soc.*, **1947**, 43, 50-67

[8] http://kinetixapp.grzybowskigroup.pl/

[9] J. Stankiewicz, L. H. Eckardt, Chembiogenesis 2005 and Systems Chemistry Workshop. *Angew. Chem. Int. Ed.*, **2006**, 45, 342 – 344

[10] J. Li, P. Nowak, S. Otto, Dynamic Combinatorial Libraries: From Exploring Molecular Recognition to Systems Chemistry. *J. Am. Chem. Soc.*, **2013**, 135, 9222-9239

[11] J. R. Nitschke, Systems chemistry: Molecular networks come of age. *Nature*, **2009**, 462, 736-738

[12] D. Newth, J. Finnigan, Emergence and Self-Organization in Chemistry and Biology. *Aust. J. Chem.*, **2006**, 59, 841-848

[13] M. Kindermann, I. Stahl, M. Reimold, W. M. Pankau, G. von Kiedrowski, Systems Chemistry: Kinetic and Computational Analysis of a Nearly Exponential Organic Replicator. *Angew. Chem.*, **2005**, 117, 6908-6913

[14] J. JP. Peyralans, S. Otto, Recent highlights in systems chemistry. *Curr. Opin. Chem. Biol.*, **2009**, 13:705–713

[15] G. von Kiedrowski, S. Otto, P. Herdewijn, Welcome Home, System Chemists!. *Journal of Systems Chemistry*, **2010**, 1:1

[16] https://jsystchem.springeropen.com/

[17] S. N. Semenov, A. S. Y. Wong, R. Martijn van der Made, S. G. J. Postma, J. Groen, H. W. H van Roekel, T. F. A. de Greef, W. T. S. Huck, Rational design of functional and tunable oscillating enzymatic networks. *Nat. Chem.*, **2015**, 7, 160-165

[18] S. N. Semenov, L. J. Kraft, A. Ainla, M. Zhao, M. Banghbanzadeh, V. E. Campbell, K. Kang, J. M. Fox, G. M. Whitesides, Autocatalytic, bistable, oscillatory networks of biologically relevant organic reactions. *Nature*, **2016**, 537, 656-660

[19] M. D. Bajczyk, P. Dittwald, A. Wołos, S. Szymkuć, B. A. Grzybowski, Discovery and Enumeration of Organic-Chemical and Biomimetic Reaction Cycles within the Network of Chemistry. *Angew. Chem. Int. Ed.*, **2018**, 57, 2367-2371

[20] R. Roszak, M. D. Bajczyk, E. P. Gajewska, R. Hołyst, B. A. Grzybowski, Propagation of oscillating chemical signals through reaction networks. *Angew. Chem. Int. Ed.*, **2019**, 58, 4520-4525

[21] https://www.rug.nl/research/centre-systems-chemistry/introduction

[22] M. B. Elowitz, S. Leibler, A synthetic oscillatory network of transcriptional regulators. *Nature*, **2000**, 403, 335-338

[23] K. Ruis-Mirazo, C. Briones, A. de la Escosura, Prebiotic Systems Chemistry: New Perspectives for the Origins of Life. *Chem. Rev.*, **2014**, 114, 285-366

[24] D. Ritson, J. D. Sutherland, Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nature Chemistry*, **2012**, 4, 895-899

[25] M. K. Runyon, B. L. Johnson-Kerner, R. F. Ismagilov, Minimal Functional Model of Hemostatis in a Biomimetic Microfluidic System. *Angew. Chem. Int. Ed.*, **2004**, 43, 1531-1536

[26] B. Helwig, B. van Sluijs, A. A. Pogodaev, S. G. J. Postma, W. T. S. Huck, Bottom-up Construction of an Adaptive Enzymatic Reaction Network. *Angew. Chem. Int. Ed.*, **2018**, 57, 14065-14069

[27] L. Zheng, H. Zhao, Y. Han, H. Qian, L. Vukovic, J. Mecinović , P. Král, W. T. S. Huck, Catalytic transport of molecular cargo using diffusive binding along a polymer track. *Nat. Chem.*, **2019**, 11, 359-366

[28] L. Stryer, J. M. Berg, J. L. Tymoczko, G. J. Gatto, Jr. (2015.) *Biochemistry: Eighth edition*. New York: Palgrave Macmillan

[29] H. A. Krebs, K. Henseleit, Studies on urea formation in the animal organism. *Hoppe-Seyler's Z. Physiol. Chem*., **1932**, 210, 33–66

[30] D. M. Greenberg, Metabolic Pathways, Academic Press, New York, 1967

[31] A. Boiteux, B. Hess, Design of glycolysis. *Philos. Trans. R. Soc. London Ser. B*, **1981**, 293, 5–22

[32] A. T. Winfree, The Prehistory of the Belousov-Zhabotinsky Oscillator. *J. Chem. Educ.*, **1984**, 61, 661-663

[33] I. E. Epstein, J. A. Pojman, O. Steinbock, Introduction: Self-organization in non-equilibrium chemical systems. *Chaos*, **2006**, 16, 037101

[34] M. A. Keller, D. Kampjut, S. A. Harrison, M. Ralser, Sulfate radicals enable a non-enzymatic Krebs cycle precursor. *Nat. Ecol. Evol.*, **2017**, 1, 1-9

[35] S. Roje, S-Adenosyl-L-metionine: Beyond the universal methyl group donor. *Phytochemistry*, **2006**, 67, 1686-1698

[36] E. Albers, Metabolic Characteristics and Importance of the Universal Methionine Salvage Pathway Recycling Methionine from 5'-Methylthioadenosine. *IUBMB Life*, **2009**, 61, 1132-1142

[37] D. M. Schmatz, The Mannitol Cycle – A New Metabolic Pathway in the Coccidia. *Parasitol. Today*, **1989**, 5, 205-208

[38] S. H. Song, C. Vielle, Recent advances in the biological production of mannitol. *Appl. Microbiol. Biotechnol.*, **2009**, 84, 55-62

[39] J. M. H. Stoop, J. D. Williamson, D. M. Pharr, Mannitol metabolism in plants: a method for coping with stress. *Trends. Plant. Sci.*, **1996**, 1, 139-144

[40] P. Naik, Essentials of Biochemistry, Jaypee Brothers Medical Publishers, London, 2011

[41] H. I. Morgan. (2007.) *The Cell Cycle: Principles of Control*, London: New Science Press

[42] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter. (2008.) *Molecular Biology of the Cell: Fifth Edition*, New York: Garland Science, Taylor & Francis Group

[43] R. F. Ludlow, S. Otto, Systems Chemistry. Chem. Soc. Rev., 2008, 37, 101-108

[44] L. E. Orgel, Molecular Replication. *Nature*, **1992**, 358, 203-2009

[45] A. Robertson, A. J. Sinclair, D. Philp, Minimal self-replicating systems. *Chem. Soc. Rev.*,**2000**, 29, 141-152

[46] T. Tjivikua, P. Ballester, J. Rebek, Jr., A Self-Replicating System. *J. Am. Chem. Soc.*, **1990**, 112, 1249-1250

[47] B. Wang, I. O. Sutherland, Self-replication in a Diels-Alder reaction. *Chem. Commun.*, **1997**, 1495-1496

[48] V. C. Allen, D. Philp, N. Spencer, Transfer of Stereochemical Information in a Minimal Self-Replicating System. *Org. Lett.*, **2001**, 3 777-780

[49] B. J. Miflin, D. Z. Habash, The role of glutamine synthetase and glutamate dehydrogenase in nitrogen assimilation and possibilities for improvement in the nitrogen utilization of crops. *J. Exp. Bot.*, **2002**, 53, 979-987

[50] C. Widmann, S. Gibson, M. B. Jarpe, G. L. Johnson. Mitogen-Activated Protein Kinase: Conservation of a Three-Kinase Module from Yeast to Human. *Physiol. Rev.*, **1999**, 79, 143-180

[51] P. Coulombe, S. Meloche, Atypical mitogen-activated protein kinases: Structure, regulation and functions. *Biochim. Biophys. Acta*, **2007**, 1773, 1376-1387

[52] B. N. Kholodenko, Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell. Biol.*, **2006**, 7, 165-176

[53] A. S. Dhillon, S. Hagan, O. Rath, W. Kolch, MAP kinase signalling pathways in cancer. *Oncogene*, **2007**, 26, 3279-3290

[54] P. J. Roberts, C. J. Der, Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene*, **2007**, 26, 3291-3310

[55] L. Chang, M. Karin, Mammalian MAP kinase signalling cascades. *Nature*, **2001**, 410, 37-40

[56] P. P. Roux, J. Blenis, ERK and p38 MAPK-Activated Protein Kinases: a Family of Protein Kinases with Diverse Biological Functions. *Microbiol. Mol. Biol. Rev.*, **2004**, 68, 320-344

[57] E. Perlson, S. Hanz, K. Ben-Yaakov, Y. Segal-Ruder, R. Seger, M. Fainzilber, Vimentin-Dependent Spatial Translocation of an Activated MAP Kinase in Injured Nerve. *Neuron*, **2005**, 45, 715-726

[58] B. N. Kholodenko, MAP kinase cascade signalling and endocytic trafficking: a marriage of convenience? *Trends. Cell. Biol.*, **2002**, 12, 173-177

[59] B. N. Kholodenko, J. F. Hancock, W. Kolch, Signalling ballet in space and time. *Nature Reviews*, **2010**, 11, 414-426

[60] C. J. Marshall, Specificity of Receptor Tyrosine Kinase Signaling: Transient versus Sustained Extracellular Signal-Regulated Kinase Activation. *Cell*, **1995**, 80, 179-185

[61] M. Cargnello, P. P. Roux, Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases. *Microbiol. Mol. Biol. Rev.*, **2011**, 75, 50-83

[62] I. R. Epstein, K. Showalter, Nonlinear Chemical Dynamics: Oscillations, Patterns and Chaos. *J. Chem. Phys.*, **1996**, 100, 13132-13147

[63] I. Prigogine, R. Lefever, Symmetry Breaking Instabilities in Dissipative Systems. II. *J. Chem. Phys.*, **1968**, 48, 1968-1700

[64] R. J. Field, H. –D. Försterling, On the Oxybromine Chemistry Rate Constants with Cerium Ions in the Field-Koros-Noyes Mechanism of the Belousov-Zhabotinskii Reaction: The Equilibrium $HBrO_2 + BrO_3^- + H^+ \leftrightarrow 2BrO_2^\bullet + H_2O$. *J. Phys. Chem.*, **1986**, 90, 5400-5407

[65] A. Sirimungkala, H. –D. Försterling, V. Dlask, Bromination Reactions Important in the Mechanism of the Belousov-Zhabotinsky System. *J. Phys. Chem. A*, **1999**, 103, 1038-1043

[66] R. J. Field, R. M. Noyes, Oscillations in chemical systems. IV. Limit cycle behaviour in a model of a real chemical reaction. *J. Chem. Phys.*, **1974**, 60, 1877-1884

[67] M. J. Boulware, J. S. Marchant, Timing in Cellular Ca^{2+} Signaling. *Curr. Biol.*, **2008**, 18, R772

[68] P. Tompa, R. Tóth-Boconádi, P. Friedrich, Frequency decoding of fast calcium oscillations by calpain. *Cell Calcium*, **2001**, 29, 161-170

[69] E. Smedler, P. Uhlen, Frequency decoding of calcium oscillations. *Biochim. Biophys. Acta*, **2014**, 1840, 964-969

[70] P. J. Cullen, P. J. Lockyer, Integration of Calcium and Ras Signalling. *Nature Reviews*, **2002**, 3, 339-348

[71] E. R. Kandell, J. H. Schwartz, T. M. Jessel, S. A. Siegelbaum, A.J. Hudspeth (2012.) *Principles of Neural Science, Fifth Edition*. McGraw-Hill Medical, New York

[72] P. I. Hanson, T. Meyer, L. Stryer, H. Schulman, Dual Role of Calmodulin in Autophosphorylation of Multifunctional CaM Kinase May Underlie Decoding of Calcium Signals. Neuron, 1994, 12, 943-

[73] M. Yi, Q. Zhao, J. Tang, C. Wang, A theoretical modeling for frequency modulation of Ca^{2+} signal on activation of MAPK cascade. *Biophys. Chem.*, **2011**, 157, 33-42

[74] S. Kupzig, S. A. Walker, P. J. Cullen, The frequencies of calcium oscillations are optimized for efficient calcium-mediated activation of Ras and the ERK/MAPK cascade. *PNAS*, **2005**, 102, 7577-7582

[75] D. A. Beard, H, Qian,. (2008.) *Chemical Biophysics: Quantitative Analysis of Cellular Systems*. New York: Cambridge University Press

[76] M. Fialkowski, K. J. M. Bishop, V. A. Chubukov, C. J. Campbell, B. A. Grzybowski, Architecture and Evolution of Organic Chemistry. *Angew. Chem. Int. Ed.*, **2005**, 44, 7263-7269

[77] K. J. M. Bishop, R. Klajn, B. A. Grzybowski, The Core and Most Useful Molecules in Organic Chemistry. *Angew. Chem. Int. Ed.*, **2006**, 45, 5548-5354

[78] https://www.reaxys.com/

[79] K. J. M. Bishop, R. Klajn, B. A. Grzybowski, The Core and Most Useful Molecules in Organic Chemistry. *Angew. Chem. Int. Ed.*, **2006**, 45, 5348-5354

[80] M. Fujita, T. Hiyama, Highly stereoselective reduction of α-substituted β-keto amides by means of hydrosilane-fluoride ion and hydrosilane-acid reagent. A practical approach to aldols of both threo and erythro configurations. *J. Am. Chem. Soc.*, **1985**, 107, 26, 8294-8296

[81] C. –J. Hou, X. –P. Hu, Sterically Hindered Chiral Ferrocenyl P,N,N-Ligands for Highly Diastereo-/Enantioselective Ir-Catalyzed Hydrogenation of α-Alkyl-β-ketoesters via Dynamic Kinetic Resolution. *Org. Lett.*, **2016**, 18, 21, 5592-5595

[82] Y. Ohtsuka, T. Oishi, An effective synthesis of medium-ring ketones. *Tetrahedron Lett.*, **1979**, 46, 4487-4490

[83] Y. Ohtsuka, T. Oishi, Medium-ring Ketone Synthesis. Synthesis of Eight- to Twelve-membered Cyclic Ketones based on the Intramolecular Cyclization of Large-membered Lactam Sulfoxides or Sulfones. *Chem. Pharm. Bull.*, **1983**, 31, 454-465

[84] T. Kawano, T. Yoshizumi, K. Hirano, T. Satoh, M. Miura, Copper-Mediated Direct Arylation of 1,3,4-Oxadiazoles and 1,2,4-Triazoles with Aryl Iodides. *Org. Lett.*, **2009**, 11, 3072-3075

[85] G. S. He, L. –S. Tan, Q. Zheng, P. N. Prasad, Multiphoton Absorbing Materials: Molecular Designs, Characterizations, and Applications. *Chem. Rev.*, **2008**, 108, 1245-1330

[86] J. Fischer, C. Kilpert, U. Klein, W. Steglich, Stereochemistry of [3.3]-sigmatropic rearrangements in the oxazole series. *Tetrahedron*, **1986**, 42, 2063-2074

[87] F. Chen, F. -F. Zhu, M. Zhang, R. -H. Liu, W. Yu, B. Han, Iminoxyl Radical-Promoted Oxycyanation and Aminocyanation of Unactivated Alkenes: Synthesis of Cyano-Featured Isoxazolines and Cyclic Nitrones. *Org. Lett.*, **2017**, 19, 3255-3258

[88] L. -M. Zhao, L. -J. Wan, S. -Q. Zhang, R. Sun, F. -Y. Ma, α-Regioselective construction of gem-bisprenyl structures via zinc-mediated prenylation of esters in THF. *Tetrahedron*, **2013**, 69, 7970-7974

[89] P. Molina, A. Ferao, P. M. Fresneda, A. Lorenzo, A. Tárraga, Synthesis of pyrido [2,1-f]-1,3,4-oxadiazine derivatives. *J. Heterocyclic Chem.*, **1983**, 20, 381-384

[90] P. Molina, P. M. Fresneda, New synthesis of pyrazole and isoxazole derivatives. *J. Heterocyclic Chem.*, **1984**, 21, 461-464

[91] P. Van der Veken, K. Senten, I. Kertesz, I. De Meester, A. M. Lambeir, M. B. Maes, S. Scharpe, A. Haemers, K. Augustyns, Fluro-Olefins as Peptidomimetic Inhibitors of Dipeptidyl Peptidases. *J. Med. Chem.*, **2005**, 48, 1768–1780

[92] H. Poleschner, M. Heydenreich, U. Schilde, Reactions of RSe-EMe_3 (E = Si, Ge, Sn, Pb) with XeF_2-RSe-F Equivalents in the Fluoroselenylation of Acetylenes. Eur. J. Inorg. Chem., **2000**, 6, 1307-1313

[93] B. K. Tivary, K. Pradhan, A. K. Nanda, R. Chakraborty, Implication of Quinazoline-4(3H)-ones in Medicinal Chemistry: A Brief Review. *J. Chem. Biol. Ther.*, **2005**, 1, 1-104

[94] T. P. Selwam, P. V. Kumar, Quinazoline Marketed Drugs – A Review. *Research in Pharmacy*, **2011**, 1, 1-12

[95] Z. Han, D. Krishnamurthy, P. Grover, H. S. Wilkinson, Q. K. Fang, X. Su, Z. –H. Lu, D. Magiera, C. H. Senanayake. A Highly Selective and Practical Method for Enantiopure Sulfoxides Utilizing Activated and Functionally Differentiated N-Sulfonyl-1,2,3-oxathiazolidine-2-oxide Derivatives. *Angew. Chem. Int. Ed.*, **2003**, 42, 2032-2035

[96] D. Ranganathan, R. Rathi, Urea cycle: chemical simulation of arginine biosynthesis. *J.Org. Chem*. **1990**, 55, 2351–2354

[97] S. Moncada, R. M. Palmer, E. A. Higgs, Nitric oxide: physiology, pathophysiology and pharmacology. *Pharmacol. Rev.*, **1991**, 43, 109-142

[98] K. Sugimoto, Y. Hoshiba, K. Tsuge, Y. Matsuya, Synthesis of Substitutes Pyrrolo[2,1-a]isoquinolines by Gold-Catalyzed Domino Cyclization of Alkynyl Iminoesters. *Synthesis*, **2016**, 48, 1855-1864

[99] I. D. Rafiqul, K. Shimada, S. Aoyagi,Y. Takikawa, C. Kabuto. Novel conversion of 6H-1,3,5-oxathiazine S-oxides into 5-membered heterocyclic compounds. *Heteroat. Chem.*, **2004**, 15, 175-186